Product Quality at the Plant Level: Plant Size, Exports, Output Prices and Input Prices in Colombia

Maurice Kugler[†]
Eric A. Verhoogen[‡]

Jan. 31, 2008

Abstract

This paper uses uniquely rich and representative data on the unit values of "outputs" (products) and inputs of Colombian manufacturing plants to draw inferences about the extent of quality differentiation at the plant level. We extend the Melitz (2003) framework to include heterogeneity of inputs and a complementarity between plant productivity and input quality in producing output quality and we show that the resulting model carries distinctive implications for two simple reduced-form correlations — between output prices and plant size and between input prices and plant size — and for how those correlations vary across sectors. We then document three plant-level facts: (1) output prices are positively correlated with plant size within industries, on average; (2) input prices are positively correlated with plant size within industries, on average; and (3) both correlations are more positive in industries with more scope for quality differentiation, as measured by the advertising and R&D intensity of U.S. firms. The correlations between export status and input and output prices are similar to those for plant size. These facts are consistent with our model of quality differentiation of both outputs and inputs, and difficult to reconcile with models that assume homogeneity or symmetry of either set of goods. Beyond recommending an amendment of the Melitz (2003) model, the results highlight shortcomings of standard methods of productivity estimation, generalize and provide an explanation for the well-known employer size-wage effect, and suggest new channels through which liberalization of trade in output markets may affect input markets and vice-versa.

[†]Wilfrid Laurier University and the Center for International Development, Harvard University, email: maurice_kugler@harvard.edu. [‡]Corresponding author: Columbia University, BREAD, CEPR, and IZA, email: eric.verhoogen@columbia.edu. Address: 420 W. 118th St. Room 1022, New York NY 10027. We would like to thank, without implicating, Marcela Eslava, Garth Frazer, Penny Goldberg, John Haltiwanger, James Harrigan, Kate Ho, David Hummels, Amit Khandelwal, Wojciech Kopczuk, Miklós Koren, Adriana Kugler, David Lee, Marc Melitz, Nathan Nunn, Esteban Rossi-Hansberg, Rafael Santos, John Sutton, Kensuke Teshima, Jim Tybout, Miguel Urquiola, David Weinstein and seminar participants at Columbia, Harvard (Kennedy School), the Firms in the Global Economy conference (Michigan), Purdue, EIIT-2007, Fordham, Syracuse, Princeton, Bar-Ilan University and the Microeconomics of Growth conference (Rio de Janeiro) for helpful conversations; Kensuke Teshima, Hitoshi Shigeoka and Juan Ovalle for excellent research assistance; and Juan Francisco Martínez, Luis Miguel Suarez, German Perez and Beatriz Ferreira of DANE for help with the plant data. Verhoogen thanks the Program for Economic Research at Columbia University for funding and the Center for Health and Well-Being at Princeton University for both funding and hospitality while this paper was being written.

1 Introduction

Recent work using information on unit values in trade-flow data has documented a number of striking empirical regularities. Schott (2004) shows that, within the most narrowly defined trade categories (7- or 10-digit, depending on the year), imports into the U.S. from richer, more capital- and skill-abundant countries have higher unit values than imports from poorer, more labor-abundant ones. Using data on bilateral trade flows at the 6-digit level, Hummels and Klenow (2005) show not only that richer countries export goods with higher unit values but also that they export a greater volume within categories. These patterns suggest quality differences within sectors, and have kindled a resurgence of research on vertical differentiation in international trade.[1] It is not clear, however, whether they reflect quality variation across individual *firms* or simply variation across sub-sectors, for instance at the (unobserved) 12-digit level. As a consequence, it is not clear what implications these results carry for the burgeoning theoretical and empirical literature on heterogeneous firms.[2] At the same time, because plant- or firm-level datasets typically lack product-level information — in particular, information on prices and physical quantities — it has been difficult to investigate the extent of quality differentiation at the firm or plant level.

In this paper, we use simple reduced-form correlations in uniquely rich and representative data on Colombian manufacturing plants to evaluate the extent of quality differentiation at the plant level. The data, from yearly censuses over the period 1982-2005, contain detailed information on the unit values and physical quantities of both "outputs" (products) and inputs, for use in constructing the national producer price index. To our knowledge, these data represent the richest source of product-level information available in a nationally representative plant-level dataset in any country.

Like Hummels and Klenow (2005), Hallak and Schott (2005) and others, we do not observe product quality directly and we must make inferences about quality from information on prices and quantities. Our first main empirical finding is that the analogue of the above Hummels-

[1] Other notable recent contributions using unit value information in trade-flow data include Hallak (2006), Hallak and Schott (2005), Choi, Hummels, and Xiang (2006), and Khandelwal (2007).

[2] For a review, see Tybout (2003).

Klenow result holds at the plant level: *on average, within narrow industries, output prices and plant size are positively correlated.* A similar pattern holds for the correlation of output prices and export status. We also note, however, that any of a variety of plant-level demand shifters could generate this positive correlation, even in the absence of quality differentiation across plants, as for instance in the framework of Foster, Haltiwanger, and Syverson (forthcoming). In other words, the implications of quality models are likely to be *observationally equivalent* to those of a number of other plausible models in trade-flow data or in plant-level data containing only output prices and quantities.

To break this observational equivalence, we push the implications of quality models on two key dimensions. First, we argue that differences in the quality of plants' outputs are likely to be accompanied by differences in the quality of their *inputs* as well. Drawing on the O-ring theory of Kremer (1993), Verhoogen (2008) hypothesized that labor quality and plant productivity are complementary in the production of output quality. Here we generalize the hypothesis to apply to material inputs as well as labor inputs, name it the *quality-complementarity hypothesis*, and extend the Melitz (2003) model to accommodate it. The resulting general-equilibrium framework nests the standard Melitz (2003) model and a quality model akin to Verhoogen (2008) as special cases.[3] In this framework, the hypothesis of complementarity between input quality and plant productivity in determining output quality — the quality-complementarity hypothesis — implies that larger plants will in general pay higher prices for inputs. Our next main empirical finding is that this prediction holds in the data: *on average, within narrow industries, input prices and plant size are positively correlated.* Again, the pattern for export status and input prices is similar

[3]In an important series of papers and books, John Sutton and Avner Shaked have developed an alternative approach that emphasizes the role of fixed and sunk costs in raising product quality (Shaked and Sutton, 1982, 1987; Sutton, 1991, 1998, 2007). Because our goal is to provide a framework in which to analyze the empirical relationship between the variable costs of producing quality — in particular, input prices — and plant size, we abstract from such fixed costs, which are not observed in the Colombian data. An attractive feature of our framework is its tractability in analyzing the quality choices of large numbers of heterogeneous firms. That said, two insights from Sutton and Shaked's work are worth emphasizing. First, the constant-elasticity-of-substitution (CES) demand framework is poorly suited to analyzing market concentration, because it is difficult to reconcile with the fact that in many industries the number of market players remains fixed as the market grows large. Second, even under alternative demand structures, fixed costs of improving quality are generally necessary in order to generate the result that the number of market players remains fixed (and that firms cannot profitably enter at the low-quality end of the product spectrum) as the market grows. (See in particular Sutton (1991, Ch. 3).)

to that for plant size. We also present evidence that market power of buyers or sellers in input markets cannot fully explain the positive input price-plant size correlation.

Second, we argue that the output price-plant size and input price-plant size correlations are likely to vary in a systematic and predictable way across industries — in particular, industries with greater scope for quality differentiation are likely to have greater output price-plant size and input price-plant size slopes. As a measure of the scope for quality differentiation at the industry level, we follow Sutton (1998) in using the ratio of expenditures on advertising and R&D to sales for U.S. firms as reported in the Line of Business Survey of the U.S. Federal Trade Commission (FTC). These data have been widely used to measure R&D and advertising intensity (e.g. Cohen and Klepper (1992), Brainard (1997), and Antras (2003)), and Sutton (1998) has given a rigorous justification for the use of the advertising and R&D intensity as a measure of the scope for quality differentiation. Our third main empirical finding is that the price-plant size correlations vary across sectors in the way our theory predicts: *the slopes of both output prices and input prices vs. plant size are increasing in the scope for quality differentiation across industries, as measured by the advertising and R&D intensity of U.S. firms.*

Our main conclusion is that the three main empirical findings — the positive average output price-plant size slope, the positive average input price-plant size slope, and the fact that both slopes are increasing in the scope for quality differentiation across industries — are consistent with our model of quality differentiation of both outputs and inputs, and are difficult to reconcile with models that impose symmetry or homogeneity of either set of goods. While the particular quality model we present uses a number of special functional forms, the cross-sectional predictions we highlight — as well as the distinctiveness of those predictions relative to other models — are likely to be robust for a broad set of demand and production specifications.

This conclusion carries a number of broader implications. First, it recommends an amendment of the influential Melitz (2003) model. Although Melitz (2003) is careful to point out that his model is consistent with the existence of quality differentiation given an appropriate choice of units (as we discuss in more detail below), much recent work has taken the model at face value and imposed the assumption that products are symmetric in the actual units observed in data. Furthermore, both the standard and the "quality" interpretations of the Melitz (2003) model impose the assumption that inputs are homogeneous. Our results suggest that both of these assumptions are inappropriate, and that heterogeneous-plant models should explicitly allow for quality differentiation of both outputs and inputs. The model we present in this paper is one example of a general-equilibrium model that satisfies this criterion.

Second, our results highlight shortcomings of widely used methods of productivity estimation. A standard approach is to deflate both output revenues and input expenditures by sector-level price indices, and to estimate productivity as the residual in a regression of log deflated output revenues on log deflated input expenditures. Katayama, Lu, and Tybout (2006) have argued that even if the coefficients of this regression can be estimated consistently, the resulting productivity estimates confound (at least) four distinct dimensions of heterogeneity across plants: (1) productive efficiency, (2) mark-ups, (3) output quality, and (4) input prices, which in part reflect input quality. They also note that the mere availability of data on physical units of inputs and outputs is not sufficient to identify productive efficiency separately from the other factors without further homogeneity assumptions. While a number of techniques have recently been developed to separate technical efficiency and mark-ups (see e.g. Melitz (2000), Bernard, Eaton, Jensen, and Kortum (2003)), comparatively little attention has been paid to the quality dimensions, especially to the heterogeneity in input quality.[4] Our results provide empirical reinforcement for the argument that ignoring heterogeneity in output and input quality is likely to lead to misleading inferences.

Third, our results generalize and provide a possible explanation for a familiar and well-established finding in labor economics: the employer size-wage effect. Labor is the one input for which both quantity (i.e. hours) and price (i.e. wage) are commonly observed in nationally representative plant-level datasets, and wages have long been known to co-vary positively with

[4]Exceptions include de Loecker (2007) and the Katayama, Lu, and Tybout (2006) paper itself, both of which structurally estimate demand systems to help distinguish the contributions of mark-ups, demand shocks (i.e. output quality) and productive efficiency. A valid alternative approach has been to focus on homogeneous industries where quality differentiation is likely to be limited (Foster, Haltiwanger, and Syverson, forthcoming; Syverson, 2004). Developing methods of productivity estimation that can be applied to vertically differentiated sectors while weakening functional-form assumptions would seem to be a fruitful direction for research.

plant size (Brown and Medoff, 1989).[5] To our knowledge, our paper is the first to present evidence from broadly representative data that this pattern generalizes to material inputs. The fact that the pattern holds for material inputs as well as labor lends weight to the argument that the size-wage correlation at least in part reflects differences in labor quality (a favored hypothesis of Brown and Medoff (1989)), and not solely institutions or contracting patterns that are specific to the labor market. Our results also point to a natural explanation for the size-input price correlation: high productivity leads plants both to grow large and to produce high-quality outputs, which in turn require high-quality inputs.

Fourth, our findings point to new channels linking changes in final-good markets to changes in input markets, which are likely to play an important role in shaping how economies react to international integration. Consider the case of a developing country integrating with a richer set of countries. If rich-country consumers impose more stringent quality demands, then as poor-country final-good producers increase exports and upgrade quality they will in turn increase quality demands on input suppliers. These pressures will affect the distribution of gains from liberalization between suppliers of different qualities of inputs, which may in turn affect the political support of different constituencies for liberalization. Verhoogen (2008) developed this argument with reference to labor inputs and wage inequality; this paper suggests that the argument can be generalized to suppliers of material inputs as well. The fact that the cross-sectional price implications of the Verhoogen (2004, 2008) model, developed in the Mexican context, generalize to material inputs and hold "out of sample" in the Colombian data, raises our confidence that the quality-complementarity hypothesis is likely to hold true more generally. Note that the hypothesis also suggests that output quality choices will respond to changes in the availability of inputs of different qualities, for instance following changes in tariffs on imported inputs; this may in turn magnify the effects of such tariff changes.

The paper is organized as follows. The next section reviews related literature. Section 3.1 develops our extension of the Melitz (2003) model to include heterogeneity of inputs and the

[5]Bernard and Jensen (1995, 1999) document that export status is positively correlated with plant size and (unsurprisingly, given the size-wage correlation) with wages.

complementarity between input quality and plant productivity. Section 3.2 discusses how the cross-sectional price implications of our model differ from those of a number of other models. Section 4 describes the dataset. Section 5 describes our simple regression method for constructing plant-level price indices and discusses our econometric strategy. Section 6 presents the results: Section 6.1 presents the baseline price-plant size correlations; Section 6.2 examines how these correlations vary across industries; and Section 6.3 investigates the extent to which market power in input markets can explain the results. Section 7 concludes.

2 Related Literature

Ours is not the first study to use information on product prices at the plant level. Several studies use data from the U.S. Census of Manufactures for a limited number of relatively homogeneous sectors for which unit values can be calculated on a consistent basis. Focusing on six homogeneous products, Roberts and Supina (1996, 2000) find a *negative* correlation between plant size and output prices. In a similar vein, Syverson (2007) and Hortacsu and Syverson (2007) find negative output price-plant size correlations in the cement and ready-mixed concrete industries, also quite homogeneous sectors. Foster, Haltiwanger, and Syverson (forthcoming) present a theoretical framework that can accommodate a positive output price-plant size correlation, as we discuss in Section 3.2.1 below, but their empirical work focuses on a small number of homogeneous industries and on the relationship between product prices and physical output productivity.[6] A subset of the studies mentioned above use information on unit values of material inputs, which are available on a consistent basis for a limited number of inputs (Dunne and Roberts, 1992; Roberts and Supina, 1996, 2000; Syverson, 2007; Hortacsu and Syverson, 2007); none of these papers explicitly reports

[6] In other work on output prices, Dunne and Roberts (1992) investigate the determinants of output prices of U.S. bread producers, but do not focus on the correlation with plant size. Abbott (1992) reports indicators of price dispersion and presents a test for product quality variation based on a perfect-competition model that yields the strong — and in our view implausible — prediction that quality differentiation will lead total revenue to be uncorrelated with observed output unit values. Unit value information on outputs (not inputs) is available in the U.S. Commodity Flow Survey (CFS) used by Hillberry and Hummels (forthcoming), but to our knowledge this information has not been related to plant size.

cross-sectional correlations of material input prices with plant size.[7] Davis, Grim, Haltiwanger, and Streitwieser (2006) analyze the prices paid by manufacturing plants for electricity, a canonical homogeneous input, and show that large purchasers pay lower prices. The primary advantage of our study over this previous work in the U.S. is that we have access to data on consistently defined output and material input prices for a much broader set of sectors — indeed, for the universe of manufacturing plants with 10 or more workers in Colombia and all the products they produce or consume. The results we report below are consistent with the U.S. findings for the most homogeneous sectors, but they also suggest that the most homogeneous sectors should not be taken as representative of the manufacturing sector as a whole.

In other related work, Aw, Batra, and Roberts (2001) use two cross-sections of plant-level data on output unit values from the Taiwanese electronics sector to investigate plant-level price differences between goods sold on the export and domestic market, but do not present evidence on cross-sectional price-plant size correlations. Bernard, Jensen, Redding, and Schott (2007) have developed a dataset based on customs declarations for international transactions that includes unit values at the plant level.[8] This exciting dataset opens up a range of new research possibilities, but it has the disadvantage that it contains unit-value information only for firms that engage in international transactions, and only for the subset of transactions that cross borders. It is not clear to what extent the results for the minority of plants that export or import in each industry can be generalized to the industry as a whole, and no price comparisons can be made between firms that engage in international transactions and firms that do not. The product-level information in the Colombian manufacturing census has been used by Eslava, Haltiwanger, Kugler, and Kugler (2004, 2005, 2006, 2007) in studies that focus on the effects of market reforms on productivity, plant turnover, and factor adjustments, rather than on price-plant size correlations or quality differentiation.

The work perhaps most closely related to ours is the independent project of Hallak and

[7] Roberts and Supina (1996, 2000) report that estimated marginal costs are declining in plant size, which they argue in part reflects lower factor prices for larger plants. Syverson (2007) reports that input prices are positively correlated with output prices, which are negatively correlated with plant size.

[8] Halpern and Koren (2007) and Eaton, Eslava, Kugler, and Tybout (2007) use similar data for Hungary and Colombia, respectively, but physical quantity information is available only in tons.

Sivadasan (2006), which is ongoing. Hallak and Sivadasan develop a model with minimum quality requirements for entering the export market and also document, in Indian data, positive output price-plant size and output price-export status correlations, similar to our first main finding above. An advantage of the Colombian data we use is that they contain information on the unit values of material inputs, which we argue is important for distinguishing the implications of quality models from a number of competing explanations.[9]

There is a large literature in empirical industrial organization using consumer- and product-level information to estimate consumers' preferences for different products or product attributes, which can be used to estimate dispersion in product quality within sectors.[10] In contrast to this literature, our approach is to take advantage of rich plant-level information to draw inferences about product quality without imposing strong functional-form assumptions on the demand side. One advantage of our approach is that we are able to relate output prices and input prices to each other and to plant size and export status, which clearly would not be possible without such data. Another advantage is that our approach is more amenable to cross-sector comparisons than approaches that rely on data on detailed product attributes.

Our work is also related to two recent papers relating unit values in trade-flow data to extensions of the Melitz (2003) model: Baldwin and Harrigan (2007) and Johnson (2007). Both papers make the provocative observation that exports to more distant or difficult-to-reach markets have higher unit values on average. This fact is difficult to explain with the standard interpretation of the Melitz (2003) model, which would suggest that more productive firms both charge lower prices and enter more distant markets than less productive firms. The fact is consistent with the hypothesis that more productive firms produce higher-quality goods and charge higher prices, a

[9]We are aware of two other independent projects using producer-level output price information that are currently ongoing. In Mexican data, Iacovone and Javorcik (2007) document that plants raise output prices in preparation for exporting, which suggests that the quality-upgrading process highlighted by Verhoogen (2004, 2008) begins prior to entry into the export market. Crozet, Head, and Mayer (2007) use price information and direct quality ratings on French wines to test the implications of a quality sorting model of trade. Neither of these projects has access to data on the unit values of material inputs.

[10]See Ackerberg, Benkard, Berry, and Pakes (2007) for a useful overview. Goldberg and Verboven (2001) use direct information on product attributes in conjunction with demand-system estimation to analyze quality-adjusted price dispersion in the European automobile market. Khandelwal (2007) uses techniques developed in this literature to estimate quality dispersion using unit-value information in trade-flow data, without relying on information on detailed product attributes.

hypothesis that is explicitly present in the Verhoogen (2004, 2008) model as well as in the variants of the Melitz model presented by Baldwin and Harrigan (2007) and Johnson (2007). The hypothesis is also implicitly present in the Melitz (2003) model itself, given a suitable redefinition of quality units — a redefinition alluded to (albeit not fully developed) in the Melitz's original paper (Melitz, 2003, p. 1699). Appendix A.1 spells out the "quality" version of the Melitz (2003) model, shows how it relates to the model we present in the next section, and shows that it is isomorphic to the Baldwin and Harrigan (2007) model, if one abstracts from differences in distance between countries.[11] As will become clear below, the key difference between our quality model and the quality Melitz/Baldwin-Harrigan model is the allowance for heterogeneity of inputs and the complementarity between plant productivity and input quality, which generates distinctive implications for input prices. An additional difference is that our framework treats product quality as a choice variable of plants, rather than a deterministic function of plants' productivity draws; this enables us to provide an account of how differences in quality distributions emerge endogenously across sectors.

3 Theory

This section presents an extension of the Melitz (2003) framework to include heterogeneity of inputs and a complementarity between input quality and plant productivity in producing output quality, along the lines of Verhoogen (2008). We also discuss verbally a number of relevant theories that do not fit within our framework. While the model we present is special in many ways, we would argue that the implications we draw are likely to be robust across a variety of specifications of demand systems and production technologies. Similarly, although our model has just two countries, one differentiated sector per country and one type of input, the implications are likely to be generalizable to a setting with many countries, sectors and inputs.

[11]The Johnson (2007) model also carries implications similar to the quality Melitz model if one abstracts from distance.

3.1 A General-Equilibrium Model with Quality-Differentiated Inputs

There are two symmetric countries and in each country there are two sectors, a differentiated, monopolistically competitive sector and a non-differentiated numeraire sector with constant returns to scale and perfect competition. In each country, a representative consumer has the following standard asymmetric CES utility function:

$$U = \left\{ \left[\int_{\lambda \in \Lambda} (q(\lambda)x(\lambda))^{\frac{\sigma-1}{\sigma}} \, d\lambda \right]^{\frac{\sigma}{\sigma-1}} \right\}^{\beta} \{z\}^{1-\beta} \tag{1}$$

where z is consumption of the non-differentiated good; λ indexes varieties (and will also index plants) in the differentiated sector; Λ represents the set of all differentiated varieties available in the market (produced in either country); σ is a parameter capturing the elasticity of substitution between varieties, where we make the standard assumption that $\sigma > 1$; $q(\lambda)$ is the quality of variety λ, assumed to be observable to all; and $x(\lambda)$ is the quantity consumed. The Cobb-Douglas form implies that the representative consumer in each country spends a constant share of income β on differentiated varieties and a share $1 - \beta$ on the non-differentiated good. Consumer optimization yields the following demand for a particular differentiated variety, λ:

$$x(\lambda) = Xq(\lambda)^{\sigma-1} \left(\frac{p(\lambda)}{P} \right)^{-\sigma} \tag{2}$$

where P is an aggregate quality-adjusted price index and X is a quality-adjusted consumption aggregate of the differentiated varieties available on the market.[12]

We assume that there is just one input, as in Melitz (2003), but we allow for heterogeneity in the quality of this input. We denote quality of the input by c, and assume that in each country there is a mass of suppliers, heterogeneous in c distributed with positive positive support

[12] Specifically,

$$P \equiv \left[\int_{\lambda \in \Lambda} \left(\frac{p(\lambda)}{q(\lambda)} \right)^{1-\sigma} d\lambda \right]^{\frac{1}{1-\sigma}} \tag{3}$$

$$X \equiv \left[\int_{\lambda \in \Lambda} (q(\lambda)x(\lambda))^{\frac{\sigma-1}{\sigma}} \, d\lambda \right]^{\frac{\sigma}{\sigma-1}} \tag{4}$$

over $(0,\infty)$. Suppliers are assumed to supply the input inelastically. We assume the marginal productivity of an input in the non-differentiated, constant-returns-to-scale numeraire sector is c and hence the price that the input commands in that sector, which we refer to as the "wage", is also c. In each country, wages in the non-differentiated sector pin down wages for the differentiated sector; plants in the differentiated sector are assumed to be price-takers in the input market.

In the differentiated sector in each country, there is a continuum of potential plants,[13] heterogeneous in what, borrowing a term from Sutton (2007), we refer to as "capability", represented by the parameter λ with distribution $g(\lambda)$ with support over $[\lambda_m, \infty)$, where $\lambda_m > 0$. We allow this capability parameter to play two roles: it may reduce unit input requirements, as productivity does in the standard interpretation of the Melitz model, or it may raise output quality for a given set of inputs.[14]

As in Verhoogen (2008), we separate the production of physical units of output and the production of quality. Each unit of output is assumed to require $\frac{1}{\lambda^a}$ units of the input, where $a \geq 0$. Also following Verhoogen (2008), we assume a complementarity between the plant's capability and the quality of the input in the production of output quality. Rather than use the Cobb-Douglas formulation of that paper, we use a CES generalization of the Cobb-Douglas form recently used in a different context by Jones (2007). It is convenient to parameterize the contribution of plant capability as λ^b where $b \geq 0$ and to parameterize the contribution of input quality as c^2.[15] The production function for quality is then:

$$q(\lambda) = \left[\frac{1}{2}\left(\lambda^b\right)^\alpha + \frac{1}{2}\left(c^2\right)^\alpha\right]^{\frac{1}{\alpha}} \tag{5}$$

[13]In this section, we treat plants as decision makers, equivalent to single-establishment firms. In the empirical section we have data only on plants and do not know which firms they belong to, and for this reason refer to plants rather than firms in the theory.

[14]Sutton (2007) uses the term "capability" to refer to a pair of parameters, one reflecting unit input requirements and the other governing quality for a given set of inputs. Here we collapse the two dimensions of plant heterogeneity to one dimension. The Hallak and Sivadasan (2006) model mentioned above maintains the two dimensions of heterogeneity. See also Brooks (2006).

[15]The choices of $\frac{1}{2}$ as the weighting factors and the quadratic form c^2 are convenient but not essential. If the quality production function (5) were instead:

$$q(\lambda) = \left[\mu\left(\lambda^b\right)^\alpha + (1-\mu)\left(c^\gamma\right)^\alpha\right]^{\frac{1}{\alpha}}$$

then the conditions $0 < \mu < 1$ and $\gamma > 1$ would be sufficient.

The parameter α reflects the degree of complementarity between capability and input quality. As α becomes more negative, the complementarity increases; as $\alpha \to -\infty$, the function becomes Leontief, where quality is the minimum of λ^b and c^2. Following Jones (2007), we assume $\alpha < 0$; this will ensure that the second-order conditions for a maximum in the plants' optimization problem are satisfied. The parameter b reflects the technological ease or difficulty in translating higher plant capability into improved product quality. If $b = 0$ then superior capability will not translate into higher quality and outputs will be symmetric across plants; a higher b reflects a greater scope for quality differentiation. A high b corresponds loosely to what Khandelwal (2007) (borrowing the term from Grossman and Helpman (1991)) calls a long "quality ladder" and more closely to what Sutton (1998) calls a high "escalation parameter" in a sector with a single technological "trajectory" (i.e. one technologically related group of products).[16] To keep the model simple, we have not introduced a parameter capturing the willingness of consumers to pay for product quality, which may also vary across sectors. We would expect such differences across sectors to play a similar role as differences in the technological possibilities for quality upgrading. That is, one could interpret a higher b as indicating either greater technological ease in improving quality or greater willingness of consumers to pay for product quality improvements, or both.

The remainder of the model is as in Melitz (2003). There is a fixed cost of production, f, and no cost of differentiation, with the result that all plants differentiate. Plants are constrained to produce one good. Plants do not know ex ante what their capability is and must pay an investment cost f_e, paid to input suppliers, in order to receive a capability draw. There is an exogenous probability of death δ in each period. There is a fixed cost of producing for the export market, f_{ex}, where $f_{ex} > f$ by assumption.[17] In the interests of simplicity, we assume that there are no variable costs of trade. We focus on steady-state equilibria in which the number of entrants equals the number of deaths and the distribution of capabilities in the market does not change.

To simplify the exposition, we first focus on plants' decisions conditional on being only in

[16] In the case of an industry with a single technological trajectory, Sutton's escalation parameter α varies inversely with the elasticity of required fixed and sunk investments (i.e. R&D and advertising expenditures) with respect to the resulting quality, which he labels β (Sutton, 1998, Ch. 3).

[17] As in Melitz (2003), it does not matter whether we think of f_{ex} as a per-period fixed cost or as the amortized per-period portion of a single, large sunk cost paid when first entering the export market.

the domestic market, and then briefly discuss the general-equilibrium analysis of the two-country economy (the details of which appear in Appendix A.2). Plants choose input quality and output price. The input quality determines the wage (in our simple model, the two are equal); input quality and λ together determine output quality, which, together with the output price, determines the number of units sold. The profit-maximization problem facing each plant, conditional on entering only the domestic market, is the following:

$$\max_{p,c}\left\{\left[p-\frac{c}{\lambda^a}\right]x-f\right\} \tag{6}$$

where x is given by (2).

Each plant in the continuum of plants is infinitesimally small and ignores the effects of its decisions on the aggregates X and P. Optimization yields the following:

$$\begin{aligned} c(\lambda) &= \lambda^{\frac{b}{2}} && \text{(7a)}\\ q(\lambda) &= \lambda^{b} && \text{(7b)}\\ p(\lambda) &= \left(\frac{\sigma}{\sigma-1}\right)(\lambda)^{\frac{b}{2}-a} && \text{(7c)}\\ r(\lambda) &= \left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1} XP^{\sigma}(\lambda)^{\eta} && \text{(7d)} \end{aligned}$$

where $r(\lambda)$ represents revenues and we define $\eta \equiv (\sigma-1)\left(\frac{b}{2}+a\right) > 0$.[18]

Several points are important to notice. First, marginal cost is $\frac{c}{\lambda^a} = (\lambda)^{\frac{b}{2}-a}$ (since the non-differentiated sector pins $w = c$) and price is a fixed mark-up over marginal cost, as is standard in models with Dixit-Stiglitz (1977) CES demand specifications.

Second, plant size, as measured by revenues, is unambiguously increasing in plant capability.

Third, if there is no scope for quality differentiation — that is, if $b = 0$ — then (as long as $a > 0$) this model reduces to the Melitz (2003) model.[19] When $b = 0$, there is no complementarity between plant capability and input quality, and all plants choose the same wage. Marginal cost is

[18] The fact that α drops out of these expressions is a consequence of the choice of the exponent on c in (5). In general, if the exponent were γ in place of 2 (see footnote 15) then $c(\lambda)$ and hence $p(\lambda)$ would depend on α.

[19] To reproduce the Melitz (2003) model to the letter, three additional minor modifications are required. See Appendix A.1.

then declining in plant capability, since capability reduces unit input requirements. Because the mark-up is constant, $p(\lambda)$ also declines in λ. In the standard interpretation of the Melitz model, $p(\lambda)$ is taken to represent observed output prices. Thus the standard interpretation predicts a *negative* correlation between output price and plant size. Because all plants choose the same wage, the model predicts zero correlation between input prices and plant size. Melitz (2003) is careful to point out that his model is consistent with quality differentiation given a suitable choice of quality units. In particular, if we interpret $p(\lambda)$ as reflecting price in quality units, the model can generate a zero or positive correlation between observed output price in physical units and plant size. (Appendix A.1 spells out this argument in detail.) But, again, since there is no complementarity between plant capability and input quality, the model predicts zero correlation between input prices and plant size. Columns 1 and 2 of Table 1 summarize the predictions of the standard and quality interpretations of the Melitz (2003) model.

Fourth, if there is some scope for quality differentiation — that is, if $b > 0$ — then the complementarity between plant capability and input quality generates positive relationships between plant capability λ and both input price w and output quality q.

Fifth, the previous point implies a tendency toward a positive correlation between output prices and λ. This tendency may be offset, however, by the input-requirement-reducing effect of capability. At sufficiently low values of b, the input-requirement-reducing effect will dominate, and output prices will be *declining* in plant capability, λ. At sufficiently high values of b, the quality-complementarity effect will dominate, and output prices will be *increasing* in λ. Using the fact that plant size in unambiguously increasing in λ, it is straightforward to show (1) that the output price-plant size slope is negative for low values of b ($0 \leq b < 2a$) and positive for high values of b ($b > 2a$); (2) that as long as $b > 0$ the input price-plant size slope is positive (although the relationship may be weak if b is close to zero); and (3) that both slopes increase with b.[20] Columns 3 and 4 of Table 1 summarize these predictions.

Sixth, if plant capability only affects quality conditional on inputs and does not reduce unit input requirements — that is, if $a = 0$ — then (as long as $b > 0$) this model reduces to a model akin

[20] Formally, using logs as we will in the empirical section below, it follows from (7a),(7c) and (7d) and that fact

to Verhoogen (2008), in which both output prices and input prices are unambiguously increasing in plant capability and hence positively correlated with plant size.[21]

Seventh, a possibly counter-intuitive prediction of our model is that in industries with scope for quality differentiation ($b > 0$), output quality is positively correlated with plant size. This may seem implausible to rich-country consumers of, for example, French wines or Swiss watches. While it may well be that this model fails to describe quality choices and the extreme high-quality end of many industries, it appears that it does capture an important characteristic of industrial sectors in countries at roughly Colombia's level of development. For instance, Verhoogen (2008) finds that larger plants in Mexico were more likely to have ISO 9000 certification, an international production standard commonly interpreted as a measure of product quality. We will also see below that the positive plant size-quality relationship is consistent with our findings in the Colombian data.

The remainder of the model works essentially as in Melitz (2003), and for that reason has been relegated to Appendix A.2. To summarize briefly, three conditions — a zero-profit condition for remaining in the domestic market, a zero-profit condition for entering the export market, and a free-entry condition that the ex ante expected present discounted value of paying the investment cost to receive a capability draw is zero — pin down the cut-off values for remaining in the domestic market, λ^*, and entering the export market, λ^*_{ex}. Since $f_{ex} > f$ by assumption, the cut-off for entering the export market is to the right of the cut-off for remaining in the domestic market: $\lambda^* < \lambda^*_{ex}$. (Refer to equation (A11) in the appendix.) The facts that total revenues of plants are equal to total income of input suppliers in the differentiated sector in each country and that

that $w = c$ that:

$$\ln p = \left(\frac{b-2a}{2\eta}\right)\ln r - \left(\frac{b-2a}{2\eta}\right)\ln\left[XP^{\sigma}\left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1}\right] - \ln\left(\frac{\sigma-1}{\sigma}\right) \tag{8}$$

$$\ln w = \frac{b}{2\eta}\ln r - \frac{b}{2\eta}\ln\left[XP^{\sigma}\left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1}\right] \tag{9}$$

If $0 < b < 2a$, then $\frac{d\ln p}{d\ln r} < 0$. If $0 < 2a < b$ then $\frac{d\ln p}{d\ln r} > 0$. $\frac{d\ln w}{d\ln r} > 0$ in both cases. It also follows that $\frac{\partial}{\partial b}\left(\frac{d\ln p}{d\ln r}\right) > 0$ and $\frac{\partial}{\partial b}\left(\frac{d\ln w}{d\ln r}\right) > 0$; both slopes are increasing in b.

[21] The Verhoogen (2008) model uses discrete-choice microfoundations for demand, rather than a CES representative consumer, uses a different specification of the quality production function, and is partial-equilibrium, but is otherwise similar to this model in the case where $a = 0$.

the representative consumer dedicates a constant share of expenditures to the differentiated good pin down the scale of the economy. The most important point of the solution of the two-country case for our empirical purposes is that, since the export cut-off is to the right of the cut-off for entry into the domestic market, export status is correlated with λ and hence we have the same predictions for the correlations of output and input prices with an indicator for export status as we have for the correlations with plant size.[22]

3.2 Discussion of Alternative Models

This sub-section discusses three additional types of models — models of plant-specific idiosyncratic demand shocks, models in which input suppliers have market power, and perfect competition models — and asks whether they can generate cross-sectional price implications similar to those of the model of the previous sub-section without appealing to differences in output and input quality across plants.

3.2.1 Idiosyncratic Demand Shocks

The approach of inferring product quality from output prices and quantities typically defines quality as any factor that shifts the demand curve for a product outward. But there are many factors that may lead to greater demand for the products of a particular plant that do not correspond to conventional notions of product quality. One example might be favorable contracts from a well-placed government procurement official. Another might be collusive agreements between particular plants not to compete head-on in particular markets. Another might be plant- or firm-specific import licenses (Mobarak and Purbasari, 2006). Although in Dixit-Stiglitz-type frameworks such shocks would typically not affect output prices since they would not affect marginal costs, in the context of other demand systems it is quite plausible that such idiosyncratic shocks

[22] Note that the symmetry across countries in this model implies that if plants enter the export market they will sell the same amount in the export market as in the domestic market. Thus the model does not predict a positive correlation of plant size and the export share of sales, conditional on exporting. Nonetheless, below we also use the export share of sales as an indicator of export status, partly for the purposes of comparison with existing results in the literature, and partly because it is not difficult to imagine extensions to our model in which the export share and plant capability would be positively related, for instance if capability reduced per-unit export costs as well as unit input requirements or if plants exported higher-quality goods with higher prices to richer consumers in foreign markets.

would lead plants both to raise prices and to increase output. For example, in the framework of Foster, Haltiwanger, and Syverson (forthcoming), which is based on a demand system similar to that of Melitz and Ottaviano (forthcoming) with endogenous mark-ups, plant-specific demand shocks unrelated to quality can have such an effect. (In their model, productivity tends to lower costs and output prices, as in the model above, so the prediction for the output price-plant size correlation is not unambiguous.) Plant-specific demand shocks unrelated to quality pose a challenge for the quality story, since they may generate a positive output price-plant size correlation even in the absence of heterogeneity in product quality. In other words, under some parameter values, the implications of the plant-specific demand shocks story and the quality story for the output price-plant size correlation are observationally equivalent.

To differentiate the implications of the competing theories, we turn to the relationship between input prices and plant size, and how the price-size correlations differ across sectors. Regarding input prices, Foster, Haltiwanger, and Syverson (forthcoming) consider the possibility that plants are also subject to idiosyncratic input price shocks, but in their model, a high input price is unambiguously bad for plants: it raises input costs and output prices, and leads to reduced output. This mechanism generates a *negative* correlation between input prices and plant size. While it is possible that a positive shock to plant-specific demand could coincide with a positive shock to plant-specific input prices, in their framework there is no explicit mechanism that would lead this to happen systematically.

An extension of the Foster et al. framework could generate a systematically positive input price-plant size correlation. Consider the possibility that plants have monopsony power in input markets and face upward-sloping supply curves for inputs. In this case, a plant-specific demand shock will generate an increase in derived demand for inputs, which will in turn tend to lead plants to pay a higher input price. This effect could offset the effect of shocks to input prices discussed in the previous paragraph, and generate a positive input price-plant size correlation overall, even in the absence of quality differentiation. This extension of the demand-shocks story carries different implications from our quality model for how the input price-plant size correlation varies across sectors, however. First, the mechanism would not lead us to expect a robustly positive input price-plant size correlation in competitive or near-competitive input sectors, where plants presumably face flat or very nearly flat supply curves. Second, the mechanism would not lead us to expect either the output price-plant size correlation or the input price-plant size correlation to increase with the scope for quality differentiation.

Note also that the prediction of a positive input price-plant size correlation relies on the presumption that large plants do not have significantly more monopsony power than smaller plants purchasing in the same input markets. If larger plants had more monopsony power than smaller plants — the analogue of the "Wal-Mart effect" for the manufacturing sector — then all else equal one would again expect a negative input price-plant size correlation.

Columns 5 and 6 of Table 1 summarize the predictions of the two versions of the demand-shocks model. The output price-plant size correlation may be positive or negative. In competitive input markets with plant-specific input-price shocks, we would expect a negative input price-plant size correlation. In input markets in which producers have market power, the input price-plant size correlation may be positive or negative.

3.2.2 Supplier Market Power

An alternative but related idea is that plants are subject to plant-specific demand shocks but that *suppliers* have market power in input markets. Within industries, plants facing positive demand shocks for their output may face lower elasticities of output demand, which may in turn lead them to be less sensitive to the prices of inputs. If suppliers have market power, they will optimally charge higher prices to these less price-sensitive producers. Halpern and Koren (2007) have recently presented a model with this feature, which they call "pricing-to-firm." This mechanism is consistent with both a positive correlation of output prices and plant size and a positive correlation of input prices and plant size. Note that the mechanism requires that suppliers have market power in input markets. Note also, as above, that the mechanism would not lead us to expect either the output price-plant size correlation or the input price-plant size correlation to increase with the scope for quality differentiation. Columns 7 and 8 of Table 1 summarize the predictions of this model: a positive output price-plant size correlation in all sectors, a positive input price-plant size correlation in input sectors in which suppliers have market power, and zero input price-plant size correlation in competitive input sectors.

3.2.3 Perfect-Competition Models

In this final theoretical sub-section, we briefly consider models with perfect competition. In our framework, if the elasticity of substitution were to grow without bound, the model would approach perfect competition and the plant with the lowest quality-adjusted cost would take over the entire market. A standard way of reconciling perfect competition with a non-degenerate distribution of plant sizes is to introduce marginal costs that increase with output, as in the span-of-control model of Lucas (1978). Under some conditions, such increasing-costs models are isomorphic to Melitz-type monopolistic-competition models with constant costs; see for instance Atkeson and Kehoe (2005). A model with such increasing costs and the assumptions (a) that producing higher quality goods requires higher-quality inputs, and (b) that plants with lower marginal cost curves have differentially lower costs of producing high quality could generate cross-sectional price implications similar to those of our monopolistic-competition quality model.

The key point, however, is that perfect-competition models cannot generate such correlations in the absence of quality differences of inputs and outputs. In the absence of such quality differences, models with perfect competition and increasing costs have the feature that plants expand output until price equals marginal cost; only plants with the same price and costs can produce the same good in equilibrium. Thus we would not expect a systematic relationship between either input prices or output prices and plant size. A possible alternative argument is that the plants observed to charge different prices are selling different products that should be thought of as belonging to different "industries". But in this view, there is no compelling economic reason for plants selling one good to be systematically larger or smaller than plants selling a different good. (One could assume that technology dictates that the optimal plant size is larger in industries with higher-priced goods, but this is assuming what needs to be proven.) In the absence of such an economic reason, there is no expectation of a systematic correlation between output or input prices among the plants that have been classified together in one industry. Column 9 of Table 1 summarizes the predictions of perfect-competition models without quality differences: zero cross-sectional correlation between plant size and either output or input prices.

4 Data

The data we use are from the *Encuesta Anual Manufacturera* (EAM) [Annual Manufacturing Survey], collected by the *Departamento Administrativo Nacional de Estadística* (DANE), the Colombian national statistical agency. The dataset can be considered a census of manufacturing plants with 10 or more workers. The plant-level data are the same that have been used, for instance, in Roberts and Tybout (1997), Clerides, Lach, and Tybout (1998), and Das, Roberts, and Tybout (2006). Complete data (including product-level information) are available for the 1982-2005 period. Data on exports and imports, as well as employment and earnings of blue-collar and white-collar workers, are available on a consistent basis only for 1982-1994. We construct two separate plant-level unbalanced panels, a 1982-2005 panel and a 1982-1994 panel. We observe approximately 4,500-5,000 plants in each year.

In conjunction with this standard plant survey, DANE also collects information on the value and physical quantity of each output and input of each plant, which is used to calculate national product price indices. The product classification scheme is based on the 4-digit International Standard Industrial Classification (ISIC) revision 2; DANE then adds four Colombia-specific digits. We observe approximately 6,000 distinct product codes in the data. A strength of these data is that DANE analysts have been careful about ensuring that units of measurement are homogeneous within product categories. In some cases new categories have been created, corresponding to the same physical product but using different units of measurement, in order to ensure homogeneity. From these data we are able to calculate unit values for every product produced and every input consumed by each plant. See Appendix B.1 for variable definitions and Appendix B.2

for details on the processing of the datasets.

Table 2 presents summary statistics on our two panels, the 1982-1994 unbalanced panel and the 1982-2005 unbalanced panel. Consistent with patterns for the U.S. documented by Bernard and Jensen (1995, 1999), exporting plants are larger, in terms of both sales and employment, and pay higher wages; also, a minority of plants export and conditional on exporting, plants derive a minority of their sales from the export market. Consistent with patterns for Taiwan (Aw and Batra, 1999) and Mexico (Verhoogen, 2008), exporting plants have a higher white-collar-to-blue-collar wage ratio. Exporting plants produce in a larger number of distinct output categories and purchase from a larger number of distinct input categories than non-exporters.[23]

5 Econometric Strategy

Our econometric goal is to estimate the cross-sectional correlations between output prices, input prices, plant size and export status, and compare them to the predictions of the various theoretical models summarized in Table 1. In order to compare price levels to plant characteristics, it is convenient first to aggregate the information on unit values at the product level to an "average" price at the plant level. This task is complicated by two difficulties. First, it is not clear how to compare prices or quantities across goods, since units are typically not comparable across products. Second, industry classifications of plants are to some extent arbitrary and a given good may be produced (or consumed) by plants in a number of different industries. Our solution is to use a simple regression method to estimate a plant-specific average price based on comparisons of a plant's price to those of other plants producing (or consuming) the same product, regardless of industry, controlling flexibly for the plant's product mix. Specifically, we run an OLS regression of product-level prices on full sets of product-year effects and plant-year effects and take the

[23]The fact that exporters produce in more distinct output categories than non-exporters is consistent with the prediction of the multi-product-firm theory of Bernard, Redding, and Schott (2006b) and the patterns documented in U.S. data by Bernard, Redding, and Schott (2006a).

coefficients on the plant-year effects as our plant-specific price indices. Our basic specification is:

$$\ln p_{ijt} = \alpha_t + \theta_{it} + \mu_{jt} + u_{ijt} \tag{10}$$

where i, j, and t index goods, plants, and years, respectively; $\ln p_{ijt}$ is the log real unit value of the good; α_t is a year-specific intercept; θ_{it} is a product-year effect; μ_{jt} is a plant-year effect; and u_{ijt} is a mean-zero disturbance. We run this regression separately for output prices and input prices, and refer to the estimates of the plant-year effects, $\widehat{\mu}_{jt}$, as the *output price index* and the *input price index*, respectively. These indices are identified by differences between the unit values of a given plant and unit values of other plants producing (or consuming) the same products; the effect of product composition is controlled for flexibly by the product-year effects.

Econometric identification of the plant-year and product-year effects in this model is not assured. Intuitively, the issue is that if in a particular year a plant only produces one product, and in that year the product is only produced by that plant, then it is not possible to identify the plant-year effect for that plant separately from the product-year effect for that product. A similar issue arises in the literature using employer-employee data to identify plant and person effects (Abowd, Kramarz, and Margolis, 1999). Generally speaking, the plant-year effects can only be uniquely identified for plants that are "connected" to other plants by a "switcher" — that is, a product produced (or consumed) simultaneously by more than one plant — in a given year. To ensure this, we find the largest such group of connected plants and drop the plants not in that connected set. This leads us to drop fewer than 5% of plant-year observations in the sample.

Once we have estimated the output and input price indices, the next step is to compare them to measures of plant size and export status. Our basic model for estimating the cross-sectional correlations is simply:

$$\mu_{jt} = X_{jt}\gamma + \delta_r + \eta_{kt} + v_{jt} \tag{11}$$

where μ_{jt} is the plant-year effect from (10); X_{jt} is a measure of plant size for plant j in year t; δ_r

and η_{kt} are region and industry-year effects, respectively;[24] and v_{jt} is a mean-zero disturbance. We estimate this model on the full unbalanced panel, including all available years in which a plant has complete data. Observations across years are likely not to be independent within plants; we allow for within-plant correlation across years and cluster errors by plant. The coefficient of interest in this regression is γ; this is the estimate we will compare to the theoretical predictions in Table 1. In some specifications, we estimate (11) with an indicator for export status or the export share of sales as the X_{jt} variable. In others, we interact X_{jt} with a sector-level measure to examine how the slope term, γ, varies across sectors. It is worth underlining the fact that the estimate of γ reflects a *correlation*, not a causal effect of plant size on plant-level average prices. Indeed, our argument is precisely that both plant size and prices are determined by unobserved heterogeneity in plant capability. Nonetheless, we argue that the estimate of γ is informative in the sense that it may help us to discriminate between competing models with contrasting predictions for the cross-sectional correlations.

A natural measure of plant size is gross output; this is in fact the standard measure of plant size used by the Colombian statistical agency, measured as total sales plus net intra-firm transfers plus net change in inventories. Measurement error in gross output is a potential concern, however; plant data from developing countries are notoriously noisy, and the Colombian data are no exception.[25] To the extent that the measurement error is classical, it may simply attenuate coefficient estimates toward zero. But non-classical measurement error is also a possibility. To address this concern, we use an alternative measure of plant size for which measurement error is likely to be less severe and, importantly, uncorrelated with reports of values and quantities of outputs and inputs: total employment. We use log total employment as an instrument for log total output in an instrumental-variables procedure; under the assumption that the measurement

[24] Note that it is not redundant to have industry-year effects in this regression, even though product-year effects were controlled for in the estimation of μ_{jt}. The reason is that there is not a perfect mapping from product categories to industries; two plants producing the same product may belong to two different industries, depending on the other products they produce.

[25] Concerns about measurement error explain why we do not simply regress prices on physical quantities at the *product* level. Unit values are calculated by dividing total value produced or consumed by quantity. Hence any measurement error that biases quantity up will bias unit value down, generating a spurious negative correlation between them.

errors in gross output and total employment are uncorrelated, the IV estimates will not be subject to the measurement-error bias.

The two-step procedure described by (10) and (11) is econometrically nearly equivalent to a one-step procedure in which log prices are regressed directly on X_{jt}. Substituting (10) into (11), we have:

$$\ln p_{ijt} = \alpha_t + \theta_{it} + X_{jt}\gamma + \delta_r + \eta_{kt} + \{u_{ijt} + v_{jt}\} \tag{12}$$

If the the disturbances, u_{ijt} and v_{jt}, are uncorrelated with the co-variates in this equation, the two-step estimator corresponding to (10)-(11) and the one-step estimator corresponding to (12) will converge asymptotically to the same estimate. As discussed above, X_{jt} is likely to be correlated with unobserved plant capability, contained in v_{jt}. As long as v_{jt} is uncorrelated with the other co-variates — in particular, θ_{it}, the product-year effects — then the two-step and one-step estimators are still expected to converge to the same value (albeit not the causal effect of X_{jt} on prices.) If θ_{it} and v_{it} are correlated, however, then the one-step and two-step estimators may not converge. Baker and Fortin (2001) contain a useful discussion of the relationship between such one-step and two-step estimators, and note that the two-step procedure may be preferred because it is consistent even in the presence of a correlation between θ_{it} and v_{jt}, since v_{jt} is absorbed by the plant-year fixed effects in the first step. Below we report estimates from both the two-step and one-step models, and show that they are qualitatively similar, although not identical.

6 Results

6.1 Baseline Estimates

Panel A of Table 3 presents estimates of equation (11) with the output price index as the dependent variable.[26] Columns 1 and 2 use log total output and log employment, respectively, as the measures

[26] The first step of our estimation procedure generates thousands of plant-specific price indices for outputs and inputs in each year, too many to report in a table. We simply note that the mean values of the output price index are -0.023 (standard error 0.005) for non-exporters and 0.106 (standard error 0.010) for exporters. The mean values of the input price index are -0.009 (standard error 0.002) for non-exporters and 0.050 (standard error 0.004) for

of plant size. Column 3 reports the instrumental-variable estimate of the coefficient on log total output using log employment as the excluded instrument.[27] The IV estimate for log total output is notably larger than the OLS estimate, consistent with the observation above that measurement error in gross output may generate attenuation bias. The coefficient on log employment is quite close to the IV estimate for total output, consistent with the hypothesis that employment is measured with less error than gross output. While the Column 1 coefficient is insignificant, the estimates in Columns 2 and 3, arguably less subject to bias than Column 1, indicate that output prices are *positively* correlated with plant size on average. The price index is in log terms and the estimates can be interpreted as elasticities: Column 2 suggests, for instance, that a 10% greater employment is associated with .13% higher output prices.

Panel B of Table 3 presents the analogous regressions with the input price index as the dependent variable. In moving from Column 1 to Columns 2 and 3, the coefficient on plant size falls. This suggests a non-classical measurement error bias.[28] But again, the arguably more reliable estimates are those in Columns 2 and 3. The important message of those columns is that input prices are *positively* correlated with plant size on average. The estimates suggest that 10% greater plant size is associated with .11-.12% higher input prices. Note that the magnitudes of the input-price estimates in Columns 2-3 of Panel B are reassuringly similar to the magnitudes of the corresponding results for output prices in Columns 2-3 of Panel A. We return to the question of whether the positive input price-plant size correlation can be explained by market power of buyers or sellers in input markets in Section 6.3 below.

To check robustness, Table 4 presents estimates using the one-step procedure, described by equation (12), for output prices (Panel A) and input prices (Panel B).[29] The point estimates

exporters. The average for all plants is normalized to zero in each year.

[27]To work around computational constraints due to the large number of fixed effects, we use an indirect least squares (ILS) estimator, estimating the first stage and reduced form in a single stacked regression (allowing for clustering of errors across equations), then dividing the relevant reduced-form coefficient by the corresponding first-stage coefficient and calculating the standard error by the delta method. We also implement this ILS estimator in the instrumental-variables models below.

[28]One possibility is the following. Suppose that a "producer" re-sells a good produced by a "supplier", reports the money paid to the supplier as input expenditure, reports sales of the good in total revenues, but does not include the number of physical units in quantity of the good produced. Then a regression of input unit value on gross output will yield a positively biased coefficient. Other measurement biases are possible.

[29]For computational reasons, we include industry fixed effects rather than industry-year fixed effects in this

for output prices in Panel A are larger than the corresponding estimates using the two-step method in Panel A of Table 3. One possible explanation, as noted above, is that the product-year indicator variables (θ_{it} in (12)) may be correlated with unobserved plant characteristics (v_{jt} in (12)). Nonetheless, it is reassuring that the qualitative story is similar: output prices are positively correlated with plant size. It is also reassuring that the one-step estimates for input prices in Panel B are essentially identical to the corresponding one-step estimates in Panel B of Table 3.

Table 5 uses the 1982-1994 panel, in which export status is observed on a consistent basis, to estimate correlations between export status and output and input prices. To simplify the presentation of results, hereafter we focus on the reduced-form regressions with log employment as the key co-variate; the IV estimates are similar. Results for the output price index are in Panel A, and for the input price index in Panel B. For comparison purposes, Column 1 of each panel presents a regression with log employment as the key co-variate, comparable to Column 2 of Table 3; the results are similar to those for the longer 1982-2005 panel. Columns 2 and 3 indicate that both output and input prices are higher among exporters. On average, exporters (i.e. plants with non-zero exports) have approximately 6% greater output prices and 4.7% greater input prices than non-exporters. Caution is warranted in interpreting the results in Columns 4 and 5, since, if one believes our theoretical framework, both employment and export status are functions of a single underlying capability parameter, λ, and are likely to be collinear. Subject to that caveat, the results in Columns 4 and 5 of Panel A indicate that being an exporter is associated with higher plant-level output prices even conditional on plant size.[30] The results in Columns 4 and 5 of Panel B are more mixed but by no means rule out a positive association of export status with input prices, even conditional on plant size.

As mentioned in the introduction, the one input for which unit values are commonly observed in plant-level datasets is labor. To compare our results for material inputs to results for employee

specification. Because we also include product-year fixed effects, this modification makes little difference for the point estimates.

[30] These results are consistent with the results of Hallak and Sivadasan (2006) in Indian data mentioned above. The fact that their theoretical model contains two dimensions of heterogeneity means that it is able to provide a coherent account of the finding of systematically higher prices among exporters conditional on plant size.

wages, Columns 1-5 of Table 6 present regressions that are similar to those in Columns 1-3, Panel B, Table 5 but with earnings of all employees, blue-collar employees, and white-collar employees, log earnings ratio and log skill ratio of white-collar to blue-collar workers, respectively, as the dependent variables. The table presents 15 separate regressions, of the dependent variable at the top of the column on the co-variate at left (as well as industry-year and region effects). We see clear evidence that the earnings of both blue-collar and white-collar workers, as well as the relative earnings of white-collar workers, are greater in larger plants and in plants with more exports. The positive wage-plant size relationship is a robust and familiar fact (Brown and Medoff, 1989), and the positive wage-exporting relationship is also consistent with long-established results (Bernard and Jensen, 1995, 1999). The positive relationships between wage inequality and plant size and between wage inequality and exporting in Column 4 are less well known, but are also consistent with findings from Taiwan (Aw and Batra, 1999) and Mexico (Verhoogen, 2008). The results in Column 5 are less robust but nonetheless are suggestive that the skill ratio is higher in larger plants and plants with greater exports.

6.2 Comparison Across Industries

The results above indicate that output and input prices are positively correlated with plant size when we constrain the slope coefficient to be the same across industries. But these average correlations may conceal significant differences across sectors. As discussed in Section 3.1 above, our model would lead us to expect negative output price-plant size and zero or low input price-plant size correlations in homogeneous industries yet strongly positive correlations in industries with more scope for quality differentiation.

As mentioned in the introduction, our measure of the scope for quality differentiation at the industry level is the ratio of total industry advertising and R&D expenditures to total industry sales for large U.S. firms from the 1975 U.S. Federal Trade Commission (FTC) Line of Business Survey.[31] The Line of Business Program, which was in existence from 1974 to 1977, is unique in that it required firms to break down advertising and R&D expenditures by industry as opposed to

[31] We are indebted to John Sutton for suggesting this measure.

reporting consolidated figures at the firm level. As a consequence, the data are generally perceived to be the most accurate industry-level information on advertising and R&D expenditures, and have been used in a large number of studies, including Cohen and Klepper (1992), Brainard (1997), Sutton (1998), and Antras (2003). In the context of a model with fixed costs of improving quality, Sutton (1998) demonstrates rigorously that there is a mapping between the (unobserved) scope for quality differentiation in an industry and the (observed) extent of fixed investments in raising quality, which we measure here by the advertising and R&D/sales ratio.[32] Although we use a different model in this paper, the same intuition carries through: if incurring greater costs to raise consumer willingness to pay is ineffective, profit-maximizing firms will not incur the costs; if such costs are observed, it must be that they are effective. Under the assumption of optimal behavior by firms, we can infer that the scope for raising consumers' willingness to pay — that is, the scope for quality differentiation — is greater in industries where firms invest more in advertising and R&D. The advertising/sales ratio may arguably be more closely tied to consumer willingness to pay than the R&D/sales ratio, so we also run separate regressions with the advertising ratio alone. We converted the information on advertising and R&D expenditures and sales from the FTC industry classification (which approximates the 1972 U.S. Standard Industrial Classification) to the ISIC revision 2 4-digit level using verbal industry descriptions. Table 7 lists the 15 ISIC 4-digit industries with the lowest and the highest advertising and R&D/sales ratios. The industries in the two groups seem to accord roughly with an intuitive ranking of less and more quality-differentiated sectors. We report basic summary statistics on these measures of differentiation (as well as the Rauch (1999) measure of horizontal differentiation, discussed below) in Panel A of Table 8.

We report the results using the differentiation measures in Table 9. Because of slippage in the concordance process, we do not have the differentiation measures for several ISIC industries, and we lose a number of plants. For comparison purposes, Columns 1 and 6 report specifications

[32] See Theorem 3.3, the remark immediately following, and footnote 12 in Sutton (1998, Ch. 3). Sutton (1998) emphasizes R&D intensity and Sutton (1991) emphasizes advertising expenditures, but the two types of quality investments play broadly similar roles in the theory, which suggests that they should be aggregated in our empirical application.

similar to Column 2 of Table 3 for the modified sample; the point estimates are not statistically different from those in the earlier table. The results in Columns 2-3 and 7-8 for the interactions of log employment with the advertising/sales ratio or the advertising and R&D/sales ratio are consistent with the predictions of our model above: the output price-plant size slope and the input price-plant size slope are significantly more positive in industries with more scope for quality differentiation.

A potential concern with the specification in Columns 2-3 and 7-8 is that the differences in slopes across sectors may be reflecting *horizontal* rather than vertical differentiation. Theoretically, one might very well expect sectors with greater horizontal differentiation to have greater price-plant size correlations: in our model, for a given level of b (as long as $b > 2a$), a greater degree of horizontal differentiation (a lower σ) will give rise to steeper output price-plant size and input price-plant size slopes.[33] In the idiosyncratic demand-shocks story as well, the positive output price-plant size correlation is likely to be stronger in more horizontally differentiated sectors. To address this issue, we control for the effect of differences in horizontal differentiation by including an interaction of plant size with a widely used measure of horizontal differentiation: the Rauch (1999) differentiation measure, based on whether a good is traded on a commodity exchange, has a quoted price in industry trade publications, or neither. (Details of the construction of the measure and conversion to the ISIC rev. 2 industry categories, which generated some fractional values, are in Appendix B.2.) The specifications with the additional Rauch interaction are reported in Columns 4-5 and 9-10. We find that more horizontally differentiated sectors have significantly greater output price-plant size slopes but not significantly greater input price-plant size slopes. The important point, however, is that the coefficient estimates for the indicators of *vertical* differentiation, the advertising/sales ratio and the advertising and R&D/sales ratio, are largely unaffected by the inclusion of the horizontal differentiation measure. The estimates for these coefficients in Columns 4-5 and 9-10 are not statistically distinguishable from those in Columns 2-3 and 6-7.

[33]Equations (8) and (9) in footnote 20 imply that $\frac{\partial}{\partial\sigma}\left(\frac{d\ln w}{d\ln r}\right) < 0$ and $\frac{\partial}{\partial\sigma}\left(\frac{d\ln w}{d\ln r}\right) < 0$ and hence that the slopes are increasing in the extent of horizontal differentiation.

A final important point about Table 9 is that the estimates for log employment without interactions in Columns 4-5 are negative and significant. That is, we find a negative output price-plant size correlation for the most homogeneous — least horizontally and least vertically differentiated — sectors. This finding is consistent both with our model and with previous findings by Roberts and Supina (1996, 2000), Syverson (2007) and Hortacsu and Syverson (2007) for homogeneous sectors in the U.S. But the estimates on the interaction terms, as well as the results for the average output price-plant size slopes (for instance in Column 1), suggest that the most homogeneous sectors are not representative of the manufacturing sector as a whole.

6.3 Evaluating Alternative Hypotheses

As discussed in Sections 3.2.1 and 3.2.2 above, the demand-shock and pricing-to-firm models can generate positive input price-plant size slopes when inputs markets are characterized by market power, either monopsony power of purchasers in the demand-shocks story or monopoly power of suppliers in the pricing-to-firm story. In this sub-section, we investigate whether this possibility can explain the positive input price-plant size correlation.

As measures of market power, we construct Herfindahl indices at the 8-digit level of inputs, separately for buyers and sellers in input markets. The Herfindahl index for purchasers is the sum of squared expenditure shares, where expenditure share refers to a given purchaser's share of total expenditures on the input. The Herfindahl index for suppliers is the sum of squared market shares, where market share refers to a given supplier's share of total sales the input. Panel B of Table 8 presents summary statistics on these measures, as well as on the raw number of purchasers and suppliers.

Consider the monopsony power of purchasers. Panel A of Table 10 presents product-level regressions for inputs as in Panel B of Table 4 but interacting log employment with the purchaser Herfindahl index. To check robustness with respect to functional form, we also transform the index into a binary variable, assigning 0 to input sectors below the median value of the index and 1 to input sectors above the median. For comparison purposes, Column 1 replicates Column 2 in Panel

B of Table 4 for the sample of input-plant-year observations for which all of the market power measures can be calculated.[34] The coefficients on the interactions in Columns 2 and 3 suggest that the input price-plant size correlation is slightly higher in input markets where purchasers are more concentrated, but the estimate is insignificant in Column 2 and only marginally significant in Column 3. The more important point is that the coefficient on (uninteracted) log employment remains positive and significant, indicating that there is still a positive estimated input price-plant size slope even in input markets that approximate perfectly competitive sectors. A possible objection to the specifications in Columns 2-3 is that the effect of purchaser monopsony power may be stronger in sectors characterized by horizontal differentiation. Columns 4-6 report regressions similar to Columns 1-3, but for input sectors with values of the Rauch (1999) measure above the median value. The results for the coefficient on uninteracted log employment remain positive and significant.

Now consider the market power of suppliers. Panel B of Table 10 presents product-level regressions similar to those in Panel A, but interacting log employment with the supplier Herfindahl index. In this case, the estimates in Columns 2-3 indicate that greater supplier monopoly power is associated with a *lower* input price-plant size slope. This is consistent with the hypothesis that the analogue of the Wal-Mart effect holds in manufacturing: large producers may be better able to bargain with suppliers with market power. Again, however, the important point is that the input price-plant size slope in the least concentrated sectors remains positive and significant. And again, focusing on more horizontally differentiated sectors in Columns 5-6 does not affect this key point. The standard error on the interaction with the supplier Herfindahl index in Column 5 is greater than in Column 2 (due in part to the smaller sample size), but the point estimates on the interaction terms in Columns 5-6 are not statistically distinguishable from those in Columns 2-3.[35]

[34] Not all of the products used as inputs by plants in the dataset are produced as outputs by plants in the dataset; hence the supplier Herfindahl index cannot be calculated for all products. For this reason the sample used in Table 10 is smaller than the sample in Table Table 4.

[35] An additional piece of evidence on the supplier market power story was reported in Table 6 above. One input in which there is scope for quality differentiation and for which suppliers arguably have little market power is unskilled labor. Union density in Colombia is low; the unionization rate in 2002 was 5.2% overall, and 4.7% in the private sector — low by Latin American standards (Farné, 2004). In Table 6 we saw that there is a strong positive

Overall, we interpret Table 10 as suggesting that market power of buyers or sellers in input markets cannot be the complete explanation for the positive input price-plant size correlation we observe, since the positive correlation exists even in the most competitive sectors.

7 Conclusion

This paper has used simple reduced-form cross-sectional correlations between output or input prices on one hand and plant size or export status on the other to draw inferences about the extent of quality differentiation across plants. The Colombian data are a uniquely rich and representative source of information on unit values of outputs and inputs at the plant level. We have three main findings. First, output prices and plant size or export status are positively correlated within narrow industries on average. Second, input prices and plant size or export status are positively correlated within narrow industries on average. Third, both patterns are stronger in industries that have more scope for quality differentiation as measured by advertising and R&D intensity of U.S. firms. We have also shown that the input price-plant size correlation cannot be fully explained by market power of either suppliers or purchasers in input markets. These findings are consistent with our general-equilibrium Melitz-type model with quality differentiation of both outputs and inputs, and difficult to reconcile with other existing models of heterogeneous firms. We take the results as supportive of the quality-complementarity hypothesis — that input quality and plant capability are complementary in determining output quality. This hypothesis carries a number of broader implications, which were discussed in the introduction. At a minimum, this paper has laid down a challenge to theories of plant heterogeneity: they should be consistent with positive correlation of both output and input prices with plant size and export status, and with the cross-sector patterns we have documented.

correlation between plant size and the wage of unskilled (as well as skilled) workers. Given the low unionization rate, it does not seem likely that individual, non-union workers have the power to set higher wages at plants they perceive to be facing less elastic demand.

References

Abbott, T. A. (1992): "Price Dispersion in U.S. Manufacturing: Implications for the Aggregation of Products and Firms," Center for Economic Studies Research Paper 92-3, March.

Abowd, J., F. Kramarz, and D. Margolis (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 67(2), 251–333.

Ackerberg, D., C. L. Benkard, S. Berry, and A. Pakes (2007): "Econometric Tools for Analyzing Market Outcomes," in *Handbook of Econometrics, vol. 6A*, ed. by J. Heckman, and E. Leamer, chap. 63, pp. 4172–4276. Elsevier, Amsterdam.

Angrist, J., and A. Krueger (1999): "Empirical Strategies in Labor Economics," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 3A. Elsevier Science.

Antras, P. (2003): "Firms, Contracts, and Trade Structure," *Quarterly Journal of Economics*, 118(4), 1375 – 1418.

Atkeson, A., and P. J. Kehoe (2005): "Modeling and Measuring Organization Capital," *Journal of Political Economy*, 113(5), 1026 – 1053.

Aw, B. Y., and G. Batra (1999): "Wages, Firm Size, and Wage Inequality: How Much Do Exports Matter?," in *Innovation, Industry Evolution, and Employment*, ed. by D. B. Audretsch, and A. R. Thurik, pp. 13–56. Cambridge University Press, Cambridge, New York and Melbourne.

Aw, B. Y., G. Batra, and M. J. Roberts (2001): "Firm Heterogeneity and Export-Domestic Price Differentials: A Study of Taiwanese Electronics Products," *Journal of International Economics*, 54(1), 149–169.

Baker, M., and N. M. Fortin (2001): "Occupational Gender Composition and Wages in Canada, 1987-1988," *Canadian Journal of Economics*, 34, 345–376.

Baldwin, R., and J. Harrigan (2007): "Zeros, Quality and Space: Trade Theory and Trade Evidence," NBER Working Paper No. 13214.

Bernard, A. B., J. Eaton, J. B. Jensen, and S. Kortum (2003): "Plants and Productivity in International Trade," *American Economic Review*, 93(4), 1268–1290.

Bernard, A. B., and J. B. Jensen (1995): "Exporters, Jobs, and Wages in U.S. Manufacturing: 1976-1987," *Brookings Papers on Economic Activity*, pp. 67–112.

——— (1999): "Exceptional Exporter Performance: Cause, Effect, or Both?," *Journal of International Economics*, 47, 1–25.

Bernard, A. B., J. B. Jensen, S. J. Redding, and P. K. Schott (2007): "Firms in International Trade," *Journal of Economic Perspectives*, 21(3), 105–130.

Bernard, A. B., S. J. Redding, and P. K. Schott (2006a): "Multi-Product Firms and Product Switching," Unpub. paper, Tuck School of Business.

——— (2006b): "Multi-Product Firms and Trade Liberalization," Unpub. paper, Tuck School of Business, Dec.

Brainard, S. L. (1997): "An Empirical Assessment of the Proximity-Concentration Trade-off between Multinational Sales and Trade," *American Economic Review*, 87(4), 520 – 544.

Brooks, E. L. (2006): "Why Don't Firms Export More? Product Quality and Colombian Plants," *Journal of Development Economics*, 80(1), 160–178.

Brown, C., and J. Medoff (1989): "The Employer Size-Wage Effect," *Journal of Political Economy*, 97(5), 1027–1059.

Choi, Y., D. Hummels, and C. Xiang (2006): "Explaining Import Variety and Quality: The Role of Income Distribution," NBER Working Paper No. 12531.

Clerides, S., S. Lach, and J. Tybout (1998): "Is Learning by Exporting Important? Micro-Dynamic

Evidence from Colombia, Mexico and Morocco," *Quarterly Journal of Economics*, 113, 903–947.

Cohen, W. M., and S. Klepper (1992): "The Anatomy of Industry R&D Intensity Distributions," *American Economic Review*, 82(4), 773 – 799.

Crozet, M., K. Head, and T. Mayer (2007): "Quality Sorting and Trade: Firm-Level Evidence for French Wine," Unpub. paper, Université de Paris I.

DANE (2004): "Metodología de la Encuesta Anual Manufacturera," Technical document, *Departamento Administrativo Nacional de Estadística*, Dec.

Das, S., M. Roberts, and J. Tybout (2006): "Market Entry Costs, Producer Heterogeneity, and Export Dynamics," Forthcoming in *Econometrica*.

Davis, S. J., C. Grim, J. Haltiwanger, and M. Streitwieser (2006): "Electricity Pricing to U.S. Manufacturing Plants, 1963-2000," Unpub. paper, Chicago GSB, May.

de Loecker, J. (2007): "Product Differentiation, Multi-Product Firms, and Estimating the Impact of Trade Liberalization on Productivity," NBER Working Paper No. 13155.

Dixit, A., and J. E. Stiglitz (1977): "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 67, 297–308.

Dunne, T., and M. Roberts (1992): "Costs, Demand, and Imperfect Competition as Determinants of Plant-Level Output Prices," in *Empirical Studies in Industrial Organization: Essays in Honor of Leonard W. Weiss*, ed. by D. Audretsch, and J. Siegfried, pp. 13–34. Kluwer Academic Publishers.

Eaton, J., M. Eslava, M. Kugler, and J. Tybout (2007): "The Margins of Entry into Export Markets: Evidence from Colombia," Unpub. paper, NYU.

Eslava, M., J. Haltiwanger, A. Kugler, and M. Kugler (2004): "The Effects of Structural Reforms on Productivity and Profitability Enhancing Reallocation: Evidence from Colombia," *Journal of Development Economics*, 75, 333–371.

——— (2005): "Factor Adjustments After Deregulation: Panel Evidence from Colombian Plants," NBER Working Paper No. 11656.

——— (2006): "Plant Turnover and Structural Reforms in Colombia," *IMF Staff Papers*, 53, 58 – 75.

——— (2007): "Trade Reforms and Market Selection: Evidence from Manufacturing Plants in Colombia," Unpub. paper, University of Maryland.

Farné, S. (2004): "Boletin del Observatorio del Mercado de Trabajo y la Seguridad Social No. 7," Universidad Externado de Colombia.

Foster, L., J. Haltiwanger, and C. Syverson (forthcoming): "Reallocation, Firm Turnover and Efficiency: Selection on Productivity or Profitability?," Forthcoming in *American Economic Review*.

Goldberg, P. K., and F. Verboven (2001): "The Evolution of Price Dispersion in the European Car Market," *Review of Economic Studies*, 68(4), 811 – 848.

Grossman, G. M., and E. Helpman (1991): "Quality Ladders in the Theory of Growth," *Review of Economic Studies*, 58, 43–61.

Hallak, J. C. (2006): "Product Quality and the Direction of Trade," *Journal of International Economics*, 68, 238–265.

Hallak, J. C., and P. Schott (2005): "Estimating Cross-Country Differences in Product Quality," Unpub. Paper, Yale University.

Hallak, J. C., and J. Sivadasan (2006): "Productivity, Quality and Exporting Behavior under Minimum Quality Requirements," Unpub. paper, University of Michigan.

Halpern, L., and M. Koren (2007): "Pricing to Firm: An Analysis of Firm- and Product-Level Import Prices," *Review of International Economics*, 15(3), 574–591.

Helpman, E., M. J. Melitz, and S. R. Yeaple (2004): "Export versus FDI with Heterogenous Firms," *American Economic Review*, 94(1), 300–316.

Hillberry, R., and D. Hummels (forthcoming): "Trade Responses to Geographic Frictions: A Decomposition Using Micro-Data," Forthcoming, *European Economic Review*.

Hortacsu, A., and C. Syverson (2007): "Cementing Relationships: Vertical Integration, Foreclosure, Productivity, and Prices," *Journal of Political Economy*, 115(2), 250–301.

Hummels, D., and P. J. Klenow (2005): "The Variety and Quality of a Nation's Exports," *American Economic Review*, 95(3), 704–723.

Iacovone, L., and B. Javorcik (2007): "Getting Ready: Preparing to Export," Unpub. paper, Oxford University.

Johnson, R. C. (2007): "Trade and Prices with Heterogeneous Firms," Unpub. paper, UC Berkeley.

Jones, C. I. (2007): "Intermediate Goods, Weak Links, and Superstars: A Theory of Economic Development," Unpub. paper, UC Berkeley.

Katayama, H., S. Lu, and J. R. Tybout (2006): "Firm-Level Productivity Studies: Illusions and a Solution," Unpub. paper, The University of Sydney, Charles River Associates and Pennsylvania State University.

Khandelwal, A. (2007): "The Long and Short (of) Quality Ladders," Unpub. paper, Yale University.

Kremer, M. (1993): "The O-Ring Theory of Economic Development," *Quarterly Journal of Economics*, 108(3), 551–575.

Lucas, R. (1978): "On the Size Distribution of Business Firms," *Bell Journal of Economics*, 9(2), 508–523.

Melitz, M. (2000): "Estimating Firm-level Productivity in Differentiated Product Industries," Unpub. paper, Harvard University.

——— (2003): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71(6), 1695–1725.

Melitz, M. J., and G. I. Ottaviano (forthcoming): "Market Size, Trade and Productivity," Forthcoming in *Review of Economic Studies*.

Mobarak, A. M., and D. P. Purbasari (2006): "Corrupt Protection for Sale to Firms: Evidence from Indonesia," Unpub. paper, University of Colorado.

Rauch, J. E. (1999): "Networks versus Markets in International Trade," *Journal of International Economics*, 48, 7–35.

Roberts, M. J., and D. Supina (1996): "Output Price, Markups, and Producer Size," *European Economic Review*, 40(3-5), 909–921.

——— (2000): "Output Price and Markup Dispersion in Micro Data: The Roles of Producer Heterogeneity and Noise," in *Advances in Applied Microeconomics, vol. 9*, ed. by M. R. Baye, pp. 1–36. Elsevier Science, JAI, Amsterdam, New York and Tokyo.

Roberts, M. J., and J. R. Tybout (1997): "The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs," *American Economic Review*, 87(4), 545–564.

Schott, P. (2004): "Across-Product versus Within-Product Specialization in International Trade," *Quarterly Journal of Economics*, 119(2), 647–678.

Shaked, A., and J. Sutton (1982): "Relaxing Price Competition through Product Differentiation," *Review of Economic Studies*, 49, 3–13.

Shaked, A., and J. Sutton (1987): "Product Differentiation and Industrial Structure," *Journal of Industrial Economics*, 36(2), 131–46.

Sutton, J. (1991): *Sunk Costs and Market Structure: Price Competition, Advertising, and the Evolution of Concentration*. MIT Press, Cambridge, Mass.

——— (1998): *Technology and Market Structure: Theory and History*. MIT Press, Cambridge Mass.

——— (2007): “Quality, Trade and the Moving Window: The Globalization Process,” *Economic Journal*, 117, F469–F498.

Syverson, C. (2004): “Market Structure and Productivity: A Concrete Example,” *Journal of Political Economy*, 112(6), 1181–1233.

——— (2007): “Prices, Spatial Competition, and Heterogeneous Producers: An Empirical Test,” *Journal of Industrial Economics*, 55(2), 197–222.

Tybout, J. (2003): “Plant- and Firm-Level Evidence on the ‘New’ Trade Theories,” in *Handbook of International Trade*, ed. by E. K. Choi, and J. Harrigan. Basil Blackwell, Oxford.

Verhoogen, E. A. (2004): “Trade, Quality Upgrading and Wage Inequality in the Mexican Manufacturing Sector: Theory and Evidence from an Exchange-Rate Shock,” Center for Labor Economics, UC Berkeley, Working Paper No. 67, January.

——— (2008): “Trade, Quality Upgrading and Wage Inequality in the Mexican Manufacturing Sector,” *Quarterly Journal of Economics*, 123(2).

A Theory Appendix

A.1 "Quality" Melitz Model

This appendix spells out a "quality" interpretation of the Melitz (2003) model, which is alluded to but not made explicit in the original paper. As mentioned in the text, our model reduces to the Melitz (2003) model when $b = 0$. To reproduce the Melitz model to the letter, we make three additional minor modifications: (1) since the Melitz (2003) model does not contain a non-differentiated sector, the share of income the representative consumer spends on the non-differentiated good must be set to zero, which corresponds to $\beta = 1$ in equation (1); (2) since inputs are assumed to be homogeneous, we assume all input suppliers have $c = 1$ and $w = 1$;[36] (3) the distribution of plant capabilities must be left as a general distribution, rather than imposing the assumption of a Pareto distribution as we do in Appendix Section A.2. Let $\varphi \equiv \lambda^a$ and express other variables in terms of φ. Then (1)-(4) and (7a)-(7d) become:

$$U = X = \left[\int_{\varphi\in\Phi} x(\varphi)^{\frac{\sigma-1}{\sigma}} d\varphi\right]^{\frac{\sigma}{\sigma-1}} \tag{A1a}$$

$$P \equiv \left[\int_{\varphi\in\Phi} p(\varphi)^{1-\sigma} d\varphi\right]^{\frac{1}{1-\sigma}} \tag{A1b}$$

$$w(\varphi) = q(\varphi) = 1 \tag{A1c}$$

$$p(\varphi) = \left(\frac{\sigma}{\sigma-1}\right)\frac{1}{\varphi} \tag{A1d}$$

$$r(\varphi) = \left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1} XP^{\sigma}\varphi^{\sigma-1} \tag{A1e}$$

$$x(\varphi) = \left(\frac{\sigma}{\sigma-1}\right)^{-\sigma} XP^{\sigma}\varphi^{\sigma} \tag{A1f}$$

which correspond exactly to the equations in Melitz (2003).

Now consider the following thought experiment, which generates the quality interpretation of the Melitz model. Suppose that the above equations refer to goods measured in quality units, which we will call "utils". Suppose further that higher-φ plants, in addition to requiring fewer units of inputs to produce one util of output, also produce goods with more utils per physical unit, where utils per physical unit are given by:

$$\tilde{q}(\varphi) = \varphi^{\epsilon} \tag{A2}$$

The existence of a relationship of this kind is alluded to in Melitz (2003, p. 1699) but not explicitly specified. This functional form is convenient but not essential to make the point. Given (A2), price and quantity in physical units are given by:

$$\tilde{p}(\varphi) = p(\varphi)\,\tilde{q}(\varphi) = \left(\frac{\sigma}{\sigma-1}\right)\varphi^{\epsilon-1} \tag{A3}$$

$$\tilde{x}(\varphi) = \frac{x(\varphi)}{\tilde{q}(\varphi)} = \left(\frac{\sigma}{\sigma-1}\right)^{-\sigma} XP^{\sigma}\varphi^{\sigma-\epsilon} \tag{A4}$$

[36] If there is no complementarity between plant capability and input quality, then all plants choose the same input quality, so it is not crucial to make this simplification.

The expression for revenues is unchanged by the redefinition of units.[37]

Several remarks are in order. First, if $\epsilon > 1$, then both output price in physical units and revenues are increasing in φ and hence are positively correlated with one another. Note also that setting $\epsilon = 1$ yields a model in which higher φ corresponds to higher quality but marginal cost and hence output price in physical units are constant, as alluded to by Melitz (2003, p. 1699).

Second, this "quality" Melitz model is isomorphic to the quality model of Baldwin and Harrigan (2007, Section 4) if one abstracts from the differences in distance between countries. Baldwin and Harrigan's parameter a represents marginal cost, which here corresponds to $\varphi^{\epsilon-1}$, and their θ corresponds to $\frac{1}{\epsilon-1}$. Their assumption that $\theta > 0$ here corresponds to the condition that $\epsilon > 1$, which guarantees that output price is increasing in φ. The value-added of the Baldwin-Harrigan model over this quality Melitz model is that it explicitly considers distance and the differential selection of higher-productivity firms into more-distant markets.

Third, the key difference between this quality Melitz model (with $\epsilon > 1$) and the quality model we present in this paper lies in the role of inputs. Here output price and marginal cost per physical unit are increasing in φ because plants are using *more units of inputs of homogeneous quality* to produce each physical unit, rather than inputs of higher quality as in our model. (That is, higher-φ plants use fewer units of inputs *per util*, but since the number of utils per physical unit increases in φ faster than input requirements decline, they use more units of inputs per physical unit.) Even if one were to introduce heterogeneity of inputs in this quality Melitz framework, in the absence of the complementarity between plant capability and input quality, there would be no systematic reason for higher-φ plants to use higher-quality inputs.

Fourth, a shortcoming of the quality Melitz/Baldwin-Harrigan framework for addressing issues of quality differentiation is that quality is a deterministic function of a plant's capability draw. There is no quality choice. Hence quality does not depend on factors such as the technological possibilities for upgrading quality or consumers' willingness to pay for such improvements, and there is no endogenous variation in the extent of quality differentiation across sectors.

A.2 Solution of Two-Country Version of Model with Quality-Differentiated Inputs

In this section of the appendix, we complete the solution of our two-country general-equilibrium model with quality differentiated inputs. The symmetry of the two economies implies that the optimal choices in (7a)-(7d) do not depend on which markets a plant has entered. The fact that profitability is monotonically increasing in λ (which follows from the fact that $r(\lambda)$ is monotonically increasing in λ) implies that in each country in equilibrium there will be a cut-off value of λ for remaining in in the domestic market, call it λ^*; plants will leave immediately after receiving their capability draw if it is below λ^*. There is also a cut-off λ^*_{ex} for entering the export market. To make it possible to solve explicitly for these cut-offs, we assume that in each country capability has a Pareto distribution such that $g(\lambda) = \frac{k\lambda_m^k}{\lambda^{k+1}}$ and the corresponding c.d.f. is $G(\lambda) = 1 - \left(\frac{\lambda_m}{\lambda}\right)^k$. As in Helpman, Melitz, and Yeaple (2004), in order to ensure that both the distribution of ca-

[37] The aggregates P and X can then be rewritten as:

$$X = \left[\int_{\varphi\in\Phi} \left(\tilde{x}(\varphi)\tilde{q}(\varphi)\right)^{\frac{\sigma-1}{\sigma}} d\varphi\right]^{\frac{\sigma}{\sigma-1}} \tag{A5}$$

$$P \equiv \left[\int_{\varphi\in\Phi} \left(\frac{\tilde{p}(\varphi)}{\tilde{q}(\varphi)}\right)^{1-\sigma} d\varphi\right]^{\frac{1}{1-\sigma}} \tag{A6}$$

pability draws and the distribution of plant revenues have finite variances, we must impose an assumption on the "shape" parameter of the Pareto distribution, in our case that $k > \eta$.

The values of the cut-offs (which, again because of symmetry, are the same in each country) are pinned down by three conditions. First, the profit of the plant on the margin between remaining in the domestic market and stopping production is zero:

$$\pi(\lambda^*) = \frac{r(\lambda^*)}{\sigma} - f = 0 \tag{A7}$$

where $r(\cdot)$ is given by (7d). Second, the additional profit of entering the export market for the plant on the margin between entering the export market and producing only for the domestic market is also zero:

$$\pi_{ex}(\lambda^*_{ex}) = \frac{r(\lambda^*_{ex})}{\sigma} - f_{ex} = 0 \tag{A8}$$

where $r(\cdot)$ is again given by (7d). Third, there is a free-entry condition: the ex ante expected present discounted value of receiving a capability draw must be equal to the investment cost required to receive the draw, such that ex ante expected profits are zero. Formally, given the steady-state probability of death, δ, and assuming there is no discounting, the condition is:

$$[1 - G(\lambda^*)] \sum_{t=0}^{\infty} (1-\delta)^t \left\{ \frac{E(r(\lambda))}{\sigma} - f \right\} + [1 - G(\lambda^*_{ex})] \sum_{t=0}^{\infty} (1-\delta)^t \left\{ \frac{E_{ex}(r(\lambda))}{\sigma} - f_{ex} \right\} - f_e = 0 \tag{A9}$$

where $E(r(\lambda))$ and $E_{ex}(r(\lambda))$ are the expected per-period revenues in the domestic and export markets, respectively, conditional on being in each market, and the terms in brackets are expected profits in the domestic and export markets, conditional on being in each market. Using (A7), (A8), and the fact that $\frac{r(\lambda)}{r(\lambda^*)} = \left(\frac{\lambda}{\lambda^*}\right)^\eta$, we have that $E(r(\lambda)) = \frac{k}{k-\eta}(\sigma f)$ and $E_{ex}(r(\lambda)) = \frac{k}{k-\eta}(\sigma f_{ex})$. Then using (A9), we can solve for the entry cut-offs:

$$\lambda^* = \lambda_m \left\{ \frac{f\eta}{f_e \delta (k - \eta)} \left[1 + \left(\frac{f}{f_{ex}} \right)^{\frac{k-\eta}{\eta}} \right] \right\}^{\frac{1}{k}} \tag{A10}$$

$$\lambda^*_{ex} = \lambda^* \left(\frac{f_{ex}}{f} \right)^{\frac{1}{\eta}} \tag{A11}$$

A particularly convenient feature of the Melitz (2003) framework which carries over to this model is that these cut-off values do not depend on the scale of the economies.

In steady state, the mass of new entrants in each country — that is, potential entrepreneurs who pay the investment cost to receive a capability draw and who have a capability above the cur-off to remain in the market — is equal to the mass of plants that die:

$$M_e \left(1 - G(\lambda^*)\right) = \delta M \tag{A12}$$

where M_e is the mass of entrepreneurs who pay the investment cost and M is the mass of firms that remain in business. Combining this equation with the free-entry condition (A9), it is straightforward to show that the total amount spent on investment, $M_e f_e$ (which is paid to input suppliers), is equal to total profits in the differentiated sector, and hence that total plant revenues are equal to total payments to input suppliers in the differentiated sector.[38] Because input suppliers with

[38]Total profit is given by the mass of plants in production times the expected profit conditional on being in the

quality c are paid a wage c, the total revenues of input suppliers in each country are simply the summation of c over all input suppliers, which we label C. The symmetry between countries ensures that total revenues of all plants based in a particular country are equal to total expenditures on varieties of the differentiated good in that country:

$$\beta C = ME(r(\lambda)) + M_{ex}E_{ex}(r(\lambda)) \tag{A14}$$

where $E(r(\lambda))$ and $E_{ex}(r(\lambda))$ are as defined above and the share parameter β is from the Cobb-Douglas utility of the representative consumer (1). Using the fact that $\frac{M_{ex}}{M} = \frac{1-G(\lambda^*_{ex})}{1-G(\lambda^*)}$, we can then solve for the mass of plants:

$$M = \frac{\beta C(k-\eta)}{k\sigma f\left[1+\left(\frac{f}{f_{ex}}\right)^{\frac{k-\eta}{\eta}}\right]} \tag{A15}$$

This completes the solution of the model.

B Data Appendix

B.1 Variable Definitions

Output unit value: Value of output of 8-digit product, divided by number of physical units of product produced. Output is sales plus net intra-firm transfers plus net increase in inventories. We also refer to the output unit value (somewhat loosely, since it represents an average) as the *output price*. In 1998 Colombian pesos.

Input unit value: Value consumed of 8-digit product, divided by number of physical units of product consumed. Consumption is purchases minus net intra-firm transfers minus net increase in inventories. We also refer to the input unit value (somewhat loosely, since it represents an average) as the *input price*. In 1998 Colombian pesos.

Output price index: Coefficient on plant-year effect in regression of log real output unit value on full sets of product-year and plant-year dummies, as described by equation (10).

Input price index: Coefficient on plant-year effect in regression of log real input unit value on full sets of product-year and plant-year dummies, as described by equation (10).

Total output: Total value of output of all products, valued at factory price. Total output is sales plus net transfers to other plants in same firm plus net increases in inventories. In billions of 1998 Colombian pesos.

Exporter: Indicator variable taking the value 1 if plant has export sales > 0, and 0 otherwise.

Export share: Export sales as a fraction of total sales.

---

market:

$$\Pi = M\left\{\left[\frac{E(r(\lambda))}{\sigma} - f\right] + \frac{1-G(\lambda^*_{ex})}{1-G(\lambda^*)}\left[\frac{E_{ex}(r(\lambda))}{\sigma} - f_{ex}\right]\right\} \tag{A13}$$

where $E(r(\lambda))$ and $E_{ex}(r(\lambda))$ are defined as above. Combining (A9), (A12), and (A13), we have $\Pi = M_e f_e$.

Average earnings: Total annual wage bill of permanent, remunerated workers, in millions of 1998 Colombian pesos, divided by total number of permanent, remunerated workers on Nov. 15 of corresponding year.

Average white-collar earnings: Annual wage bill of permanent, remunerated white-collar workers, in millions of 1998 Colombian pesos, divided by number of permanent, remunerated white-collar workers on Nov. 15 of corresponding year. White-collar workers defined as managers (*directivos*), non-production salaried workers (*empleados*), and technical employees (*técnicos*). White-collar/blue-collar distinction available on a consistent basis only for 1982-1994.

Average blue-collar earnings: Annual wage bill of permanent, remunerated blue-collar workers, in millions of 1998 Colombian pesos, divided by number of permanent, remunerated blue-collar workers on Nov. 15 of corresponding year. Blue-collar workers are defined as operators (*obreros* and *operarios*) and apprentices (*aprendices*). White-collar/blue-collar distinction available on a consistent basis only for 1982-1994.

Advertising/sales ratio: Ratio of advertising expenditures to total sales at sector level, from the U.S. Federal Trade Commission (FTC) 1975 Line of Business Survey. Converted from FTC 4-digit industry classification to ISIC 4-digit rev. 2 classification using verbal industry descriptions.

Advertising + R&D/sales ratio: Ratio of advertising plus research and development (R&D) expenditures to total sales, from the U.S. Federal Trade Commission (FTC) 1975 Line of Business Survey. Converted from FTC 4-digit industry classification to ISIC 4-digit rev. 2 classification using verbal industry descriptions.

Rauch (1999) measure of differentiation: SITC 4-digit sectors classified by Rauch's "liberal" classification as "homogeneous" or "reference-priced" are assigned 0, others are assigned 1. SITC 4-digit industries were then converted to ISIC rev. 2 4-digit using concordance from OECD, which generated some fractional values.

Herfindahl index (of purchasers): Sum of squares of expenditure shares of purchasers of the corresponding 8-digit input, where the expenditure sure is the expenditure by a given purchaser as a share of total expenditures on the good.

Number of purchasers of input: Number of plants in census reporting non-zero expenditures on the corresponding 8-digit input.

Herfindahl index (of suppliers): Sum of squares of market shares of producers of the corresponding 8-digit input.

Number of producers of input: Number of plants in census reporting non-zero output of the corresponding 8-digit input.

All monetary variables have been deflated to constant 1998 values using the national producer price index. Average 1998 exchange rate: 1,546 pesos/US$1.

B.2 Data Processing

The data we use in this paper are from the *Encuesta Anual Manufacturera* (EAM) [Annual Manufacturing Survey]. Plant-level data are available over the 1977-2005 period, but product-level data are available only for 1982-2005. The EAM is a census of all manufacturing plants in Colombia with 10 or more workers, with the following qualification. Prior to 1992, the sole criterion for initial inclusion of a plant in the census was that the plant have 10 or more workers.[39] Beginning in 1992, an additional criterion was added: a plant would be included if it had 10 or more workers *or* nominal value of total output (defined as in Appendix B.1) in excess of 65 million Colombian pesos (approx. US$95,000) (DANE, 2004, p. 8). The monetary limit has been raised in nominal terms over time. There are two exceptions to these rules. First, once a plant is included in the sample it is followed over time until it goes out of business, regardless of whether the criteria for inclusion continue to be satisfied. Second, multi-plant firms are included, even if not all plants satisfy one of the above criteria. To maintain consistency of the sample over time, we removed all plants with fewer than 10 workers.

The longitudinal links between plant-level observations we use are those that are reported directly by DANE. In 1991 and again in 1992, plant identification numbers were changed, with the result that it was no longer possible to follow some plants over time, despite the fact that they remained in the dataset.[40]

From 1982-2000, the product-level data were reported using an 8-digit classification system with four digits from the International Standard Industrial Classification (ISIC) revision 2 and four Colombia-specific digits (one of which is only used for verification purposes).[41] In 2001, a new classification was constructed, with the first five digits based on the U.N. Central Product Classification (CPC) version 1.0 and two Colombia-specific digits. We used a concordance provided by DANE to convert back to the earlier product classification. There are approximately 6,000 distinct product categories. To construct a plant's 5 digit industry, we followed DANE's procedure and aggregated within plants from the 8-digit to the 5-digit level, then chose the 5-digit category with the greatest value. Our procedure differed from the DANE procedure in that we aggregated across years in order to maintain stable industry definitions, while DANE defines industry year-by-year.

In the process of data cleaning, we dropped any plant-year observation for which a key variable — total output, employment, white-collar wage, blue-collar wage or average wage — had changed by more than a factor of 5 from the previous period. To reduce the influence of outliers, we followed a suggestion of Angrist and Krueger (1999) and "winsorized" the data within each year, setting all values below the 1^{st} percentile to the value at the 1^{st} percentile, and all values above the 99^{th} percentile to the 99^{th} percentile. At the plant level, we performed this procedure for total output, employment, white-collar wage, blue-collar wage or average wage. At the product level, because of the small number of observations for many products, we winsorized real unit values within product for all years together.

We dropped product-level observations in the "not elsewhere classified" product category. The product-level data also contain an identifier (an additional digit in the product code) to indicate whether the good is produced or purchased under a sub-contracting arrangement. Goods produced under subcontract are included in total output, but we did not use the unit value information

[39] This was the sole criterion over the 1970-1992 period. Prior to 1970, an additional output criterion had been in place.

[40] Eslava, Haltiwanger, Kugler, and Kugler (2004) construct some links probabilistically (see the data appendix of that paper); we use only the links constructed on the basis of name, address and telephone information.

[41] The Spanish acronym for this classification system is CIIU2AC, for *Clasificación Internacional Industrial Uniforme revisión 2 adaptada para Colombia* [ISIC revision 2 adapted for Colombia].

for subcontracted goods, since the reported value typically does not reflect the market price. We dropped plants that were reported to be cooperatives, publicly owned, or owned by a religious organization. We also dropped plants that were missing a key variable — employment, average wage, total output — or product-level information on outputs and inputs.

We refer to the unbalanced panel consisting of all plant-year observations with complete data on total output, employment, average earnings, and outputs and inputs as the 1982-2005 panel. This panel is used in Column 4 of Table 2 and in Table 3. We also selected two plant-level subsets. First, we selected the subset of plant-year observations with complete information on white-collar and blue-collar wages and employment and on exports as a share of revenues and imported inputs as a share of expenditures. This information is available on a consistent basis only for the 1982-1994 period.[42] We refer to this subset as the 1982-1994 panel; it is used in Columns 1-3 of Table 2 and in Tables 5 and 6. Second, we selected the subset of plant-year observations for which complete sector-level information on our measures of the scope for quality differention — advertising and R&D intensity — and the Rauch (1999) measure was available. This subset is used in Table 7, Panel A of Table 8, and Table 9.

The product-level datasets containing outputs and inputs corresponding to the full 1982-2005 unbalanced panel are used in Table 4, Panel B of Table 8, and Table 10. For computational reasons, we treat observations in product categories with fewer than 24 plant-product-year observations as belonging to a single product category for the product-level analysis.

The primary sub-national administrative region in Colombia is the *departamento*, of which there are 32 plus the federal district of Bogotá. Four *departamentos* have zero plants in our sample. Another eight little-populated *departamentos* — Amazonas, Arauca, Caqueta, Casanaré, Chocó, La Guajira, Putumayo, and San Andres — together have just 184 plant-year observations in the entire 1982-2005 panel. We aggregated these eight *departamentos* into a single region.

[42]Information on exports and imported inputs is also available in 2000-2005, but the information is collected in a different way and there appear to be incomparabilities between the 1982-1994 and 2000-2005 values.

Table 1:
Heterogeneous-plant models: predicted within-industry price-plant size correlations

	Standard Melitz model	Quality Melitz model	Quality-differentiated inputs model		Plant-specific demand shocks models		Pricing-to-firm model		Perfect competition (without quality) models
			low scope for quality differentiation	high scope for quality differentiation	competitive input markets	producer monopsony power	competitive input markets	supplier monopoly power	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
output prices vs. plant size	–	+ or –	–	+	+ or –	+ or –	+	+	0
input prices vs. plant size	0	0	0 or +	+	–	+ or –	0	+	0

Table 2:
Summary statistics by export status

	1982-1994 panel			1982-2005 panel
	non-exporters	exporters	all plants	all plants
	(1)	(2)	(3)	(4)
Output	2.78	12.05	4.37	5.48
	(0.04)	(0.19)	(0.05)	(0.04)
Employment	56.69	193.81	80.17	83.22
	(0.40)	(2.06)	(0.53)	(0.40)
Avg. earnings	3.26	4.67	3.50	4.40
	(0.01)	(0.02)	(0.01)	(0.01)
White-collar earnings	4.37	6.63	4.75	
	(0.01)	(0.03)	(0.01)	
Blue-collar earnings	2.77	3.48	2.89	
	(0.00)	(0.01)	(0.00)	
White-collar/blue-collar earnings ratio	1.62	1.97	1.68	
	(0.00)	(0.01)	(0.00)	
White-collar employment share	0.29	0.33	0.30	
	(0.00)	(0.00)	(0.00)	
Number of output categories	3.48	4.53	3.66	3.66
	(0.01)	(0.04)	(0.01)	(0.01)
Number of input categories	10.42	17.26	11.59	11.81
	(0.03)	(0.15)	(0.04)	(0.03)
Export share of sales		0.17		
		(0.00)		
Import share of input expenditures	0.06	0.23	0.09	
	(0.00)	(0.00)	(0.00)	
N	49671	10259	59930	114952

Notes: Standard errors of means in parentheses. Exporter defined as export sales > 0. Export share is fraction of total sales derived from exports. Annual sales measured in billions of 1998 Colombian pesos. Annual earnings measured in millions of 1998 pesos. Average 1998 exchange rate: 1,546 pesos/US$1. Number of output or input categories refers to number of distinct categories in which non-zero sales or expenditures are reported. See Appendix B.1 for more detailed variable descriptions and Appendix B.2 for details of data processing.

Table 3:
Price indices vs. plant size, 1982-2005 panel

	OLS	Reduced form	IV
	(1)	(2)	(3)
A. Dependent variable: output price index			
log total output	0.007		0.012**
	(0.005)		(0.006)
log employment		0.013**	
		(0.006)	
industry-year effects	Y	Y	Y
region effects	Y	Y	Y
R^2	0.57	0.57	
N	114952	114952	114952
B. Dependent variable: input price index			
log total output	0.016***		0.011***
	(0.002)		(0.003)
log employment		0.012***	
		(0.003)	
industry-year effects	Y	Y	Y
region effects	Y	Y	Y
R^2	0.36	0.36	
N	114952	114952	114952

Notes: Output (input) price index defined as coefficient on plant-year effect from product-level regression of log real output (input) unit values on full sets of plant-year and product-year effects. (Refer to equation (10) in Section 5 of text.) Total output is total value of production, defined as sales plus net transfers plus net change in inventories. In Column 3, log employment is instrument for log total output; the coefficient on log employment, its robust standard error and the R^2 in the first stage are 1.075, 0.008 and 0.800, respectively. See Appendix B.1 for more detailed variable descriptions and Appendix B.2 for details of data processing. Errors clustered at plant level. Robust standard errors in parentheses. *10% level, **5% level, ***1% level.

Table 4:
Product-level prices vs. plant size

	OLS	Reduced form	IV
	(1)	(2)	(3)
A. Dependent variable: log real output unit value			
log total output	0.013**		0.017***
	(0.006)		(0.006)
log employment		0.018***	
		(0.007)	
product-year effects	Y	Y	Y
industry effects	Y	Y	Y
region effects	Y	Y	Y
R2	0.86	0.86	
N	412733	412733	412733
B. Dependent variable: log real input unit value			
log total output	0.015***		0.011***
	(0.003)		(0.003)
log employment		0.012***	
		(0.003)	
product-year effects	Y	Y	Y
industry effects	Y	Y	Y
region effects	Y	Y	Y
R2	0.74	0.74	
N	1353579	1353579	1353579

Notes: Estimates correspond to one-step method described by equation (12) in Section 5 of text. Total output is total value of production, defined as sales plus net transfers plus net change in inventories. In Column 3, log employment is instrument for log total output; the coefficient on log employment, its robust standard error and the R^2 in the first stage are 1.059, 0.010 and 0.847 in Panel A and 1.081, 0.010 and 0.838 in Panel B, respectively. Product-year and industry effects are not perfectly collinear because industry is defined as the industry category with the greatest share of plant sales, and two plants producing the same product may be in different industries. See Appendix B.1 for more detailed variable descriptions and Appendix B.2 for details of data processing. Errors clustered at plant level. Robust standard errors in brackets. *10% level, **5% level, ***1% level.

Table 5:
Price indices vs. plant size and exporting variables, 1982-1994 panel

	(1)	(2)	(3)	(4)	(5)
A. Dependent variable: output price index					
log employment	0.013*			0.005	0.011
	(0.008)			(0.008)	(0.008)
exporter		0.061***		0.056***	
		(0.020)		(0.022)	
export share			0.148**		0.130*
			(0.069)		(0.070)
industry-year effects	Y	Y	Y	Y	Y
region effects	Y	Y	Y	Y	Y
R^2	0.56	0.56	0.56	0.56	0.56
N	59930	59930	59930	59930	59930
B. Dependent variable: input price index					
log employment	0.013***			0.008**	0.013***
	(0.004)			(0.004)	(0.004)
exporter		0.046***		0.037***	
		(0.009)		(0.009)	
export share			0.050*		0.028
			(0.026)		(0.026)
industry-year effects	Y	Y	Y	Y	Y
region effects	Y	Y	Y	Y	Y
R^2	0.36	0.36	0.36	0.36	0.36
N	59930	59930	59930	59930	59930

Notes: Output (input) price index defined as coefficient on plant-year effect from product-level regression of log real output (input) unit values on full sets of plant-year and product-year effects. (Refer to equation (10) in Section 5 of text.) Exporter equals 1 if plant has exports>0, and 0 otherwise. Export share is fraction of total sales derived from exports. See Appendix B.1 for more detailed variable descriptions and Appendix B.2 for details of data processing. Errors clustered at plant level. Robust standard errors in parentheses. *10% level, **5% level, ***1% level.

Table 6:
Wage variables vs. plant size and exporting variables, 1982-1994 panel

	log avg. earnings (1)	log blue-collar earnings (2)	log white-collar earnings (3)	log earnings ratio (4)	white-collar share (5)
log employment	0.141***	0.101***	0.200***	0.098***	0.008***
	(0.003)	(0.003)	(0.004)	(0.003)	(0.002)
exporter	0.278***	0.188***	0.338***	0.150***	0.037***
	(0.009)	(0.008)	(0.011)	(0.008)	(0.004)
export share	0.322***	0.214***	0.490***	0.276***	-0.013
	(0.028)	(0.023)	(0.034)	(0.026)	(0.011)
industry-year effects	Y	Y	Y	Y	Y
region effects	Y	Y	Y	Y	Y
N	59930	59930	59930	59930	59930

Notes: Table reports 15 separate regressions, all with N=59930 and including region and industry-year effects. The coefficient on log employment, its robust standard error and the R^2 in the first stage are 1.202, 0.007 and 0.862 respectively. Exporter equals 1 if plant has exports > 0, and 0 otherwise. Export share is fraction of total sales derived from exports. See Appendix B.1 for more detailed variable descriptions and Appendix B.2 for details of data processing. Errors clustered at plant level. Robust standard errors in parentheses. *10% level, **5% level, ***1% level.

Table 7:
Industries with least and most scope for quality differentiation, as measured by advertising + R&D/sales ratio

ISIC category	Industry description	advertising ratio	adv. + R&D ratio
A. Lowest advertising + R&D/sales ratio			
3692	Cement, lime and plaster	0.000	0.002
3231	Tanneries and leather finishing	0.000	0.002
3841	Ship building and repairing	0.002	0.003
3118	Sugar factories and refineries	0.002	0.004
3530	Petroleum refineries	0.002	0.004
3311	Sawmills, planing and other wood mills	0.002	0.005
3412	Containers and boxes of paper and paperboard	0.001	0.005
3710	Iron and steel basic industries	0.001	0.006
3111	Slaughtering, preparing and preserving meat	0.005	0.006
3411	Pulp, paper and paperboard	0.003	0.008
3720	Non-ferrous metal basic industries	0.002	0.011
3691	Structural clay products	0.003	0.011
3115	Vegetable and animal oils and fats	0.010	0.013
3112	Dairy products	0.011	0.013
3842	Railroad equipment	0.001	0.014
B. Highest advertising + R&D/sales ratio			
3513	Synthetic resins, plastic materials and man-made fibres	0.007	0.045
3521	Paints, varnishes and laquers	0.015	0.045
3620	Glass and glass products	0.008	0.046
3853	Watches and clocks	0.035	0.046
3901	Jewellery and related articles	0.046	0.049
3122	Prepared animal feeds	0.042	0.050
3851	Professional and scientific equipment	0.012	0.051
3832	Radio, television and communication equipment	0.009	0.053
3116	Grain mill products	0.052	0.058
3140	Tobacco manufactures	0.076	0.082
3825	Office, computing and accounting machinery	0.007	0.085
3852	Photographic and optical goods	0.024	0.095
3131	Distilling, rectifying and blending spirits	0.119	0.121
3523	Soaps, perfumes, cosmetics and other toiletries	0.105	0.124
3522	Drugs and medicines	0.079	0.166

Notes: Data on ratio of industry advertising and R&D expenditures to total industry sales are from the U.S. Federal Trade Commission (FTC) 1975 Line of Business Survey, converted from FTC 4-digit industry classification to ISIC 4-digit rev. 2 classification using verbal industry descriptions.

Table 8:
Summary statistics, measures of differentiation and market power

	mean	median	std. dev.	min.	max.
A. Defined at level of industry (4-digit)					
advertising/sales ratio	0.019	0.010	0.024	0.000	0.119
advertising + R&D/sales ratio	0.035	0.025	0.032	0.002	0.166
Rauch (1999) measure	0.690	0.970	0.385	0	1
B. Defined at level of input (8-digit)					
herfindahl index (of purchasers)	0.62	0.58	0.31	0.01	1
number of purchasers	7.32	3	16.99	1	407
herfindahl index (of suppliers)	0.38	0.30	0.29	0.01	1
number of suppliers	47.69	14	108.12	1	1627

Notes: Data on ratio of industry advertising and R&D expenditures to total industry sales are from the U.S. Federal Trade Commission (FTC) 1975 Line of Business Survey, converted from FTC 4-digit industry classification to ISIC 4-digit rev. 2 classification using verbal industry descriptions. At SITC 4-digit level, Rauch (1999) measure set to 0 if good is "homogeneous" or "reference-priced" according to the Rauch "liberal definition", to 1 if reported not to be in either category, and then concorded to ISIC rev. 2 4-digit categories. Herfindahl index of purchasers is sum of squared expenditure shares of purchasers of input. Number of purchasers is number of plants in census reporting non-zero expenditures on input in entire census. Herfindahl index of suppliers is sum of squared market shares of producers of input. Number of producers of input is number of plants in census reporting non-zero sales of the good as an output. Averages assign equal weight to each industry (Panel A) or product-year (Panel B). See Appendix B.1 for more detailed variable descriptions and Appendix B.2 for details of data processing.

Table 9:
Price indices vs. plant size interacted with advertising + R&D intensity, 1982-2005 panel

	dep. var.: output price index					dep. var.: input price index				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
log employment	0.012*	-0.004	-0.007	-0.041***	-0.042***	0.012***	0.005	0.002	-0.001	-0.003
	(0.007)	(0.009)	(0.010)	(0.015)	(0.015)	(0.003)	(0.005)	(0.005)	(0.008)	(0.008)
log emp.*advertising ratio		0.818***		0.790**			0.354**		0.350**	
		(0.316)		(0.315)			(0.147)		(0.147)	
log emp.*(adv. + R&D) ratio			0.646**		0.591**			0.327***		0.318**
			(0.274)		(0.271)			(0.127)		(0.126)
log emp.*Rauch measure				0.051***	0.049***				0.009	0.008
				(0.016)	(0.016)				(0.009)	(0.009)
industry-year effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
region effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R^2	0.57	0.57	0.57	0.57	0.57	0.38	0.38	0.38	0.38	0.38
N	90129	90129	90129	90129	90129	90129	90129	90129	90129	90129

Notes: Output (input) price index defined as coefficient on plant-year effect from product-level regression of log real output (input) unit values on full sets of plant-year and product-year effects. (Refer to equation (10) in Section 5 of text.) Data on advertising and R&D expenditures as a share of total industry sales are from the U.S. Federal Trade Commission (FTC) 1975 Line of Business Survey, converted from FTC 4-digit industry classification to ISIC 4-digit rev. 2 classification using verbal industry descriptions. At SITC 4-digit level, Rauch (1999) measure set to 0 if good is "homogeneous" or "reference-priced" according to the Rauch "liberal definition", to 1 if reported not to be in either category, and then concorded to ISIC rev. 2 4-digit categories. Sample includes sectors for which advertising and R&D intensity and Rauch measures could be constructed. Columns 1 and 6 correspond to Column 2 of Table 3 but use reduced sample. See Appendix B.1 for more detailed variable descriptions and Appendix B.2 for details of data processing. Errors clustered at plant level. Robust standard errors in brackets. *10% level, **5% level, ***1% level.

Table 10:
Product-level input prices vs. plant size interacted with indicators for market power in input markets

	dependent variable: log real input unit value					
	all sectors			sectors with horizontal differentiation > median		
	(1)	(2)	(3)	(4)	(5)	(6)
A. Measures of purchaser monopsony power						
log employment	0.011***	0.010***	0.008**	0.018***	0.020***	0.016***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
log emp.*herfindahl index (purchasers)		0.006			-0.013	
		(0.015)			(0.024)	
log emp.*1(herfindahl index > median)			0.007*			0.003
			(0.004)			(0.006)
B. Measures of supplier monopoly power						
log employment	0.011***	0.018***	0.017***	0.018***	0.023***	0.023***
	(0.003)	(0.004)	(0.004)	(0.004)	(0.006)	(0.004)
log emp.*herfindahl index (suppliers)		-0.014**			-0.012	
		(0.006)			(0.011)	
log emp.*1(herfindahl index > median)			-0.012***			-0.013**
			(0.004)			(0.006)
product-year effects	Y	Y	Y	Y	Y	Y
industry effects	Y	Y	Y	Y	Y	Y
region effects	Y	Y	Y	Y	Y	Y
N	1143626	1143626	1143626	588982	588982	588982

Notes: Table reports 12 separate regressions, each including region and industry-year effects, with the N indicated at the bottom of the column. Herfindahl index of purchasers is sum of squared expenditure shares of purchasers of input. Number of purchasers is number of plants in census reporting non-zero expenditures on input in entire census. Herfindahl index of suppliers is sum of squared market shares of producers of input. Columns 4-6 include inputs in industries with value of Rauch (1999) measure above median. See Appendix B.1 for more detailed variable descriptions and Appendix B.2 for details of data processing. Errors clustered at plant level. Robust standard errors in parentheses. *10% level, **5% level, ***1% level.